

Rec'd PCT/PTO 03 DEC 2004

## Methods of Diagnosing & Treating Diabetes and Insulin Resistance

### CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of U.S. provisional application no. 5 60/386,429, filed June 6, 2002; U.S. provisional application no. 60/386,551, filed June 5, 2002; U.S. provisional application no. 60/387,301, filed June 7, 2002; U.S. provisional application no. 60/386,527, filed June 5, 2002; U.S. provisional application no. 60/386,521, filed June 5, 2002; U.S. provisional application no. 60/386,954, filed June 6, 2002; and U.S. provisional application no. 60/386,936, filed June 6, 2002, each of which applications is 10 herein incorporated by referenced.

### BACKGROUND OF THE INVENTION

[0002] Diabetes mellitus can be divided into two clinical syndromes, Type 1 and Type 2 diabetes mellitus. Type 1, or insulin-dependent diabetes mellitus (IDDM), is a chronic autoimmune disease characterized by the extensive loss of beta cells in the pancreatic Islets 15 of Langerhans, which produce insulin. As these cells are progressively destroyed, the amount of secreted insulin decreases, eventually leading to hyperglycemia (abnormally high level of glucose in the blood) when the amount of secreted insulin drops below the level required for euglycemia (normal blood glucose level). Although the exact trigger for this immune response is not known, patients with IDDM have high levels of antibodies against proteins 20 expressed in pancreatic beta cells. However, not all patients with high levels of these antibodies develop IDDM.

[0003] Type 2 diabetes (also referred to as non-insulin dependent diabetes mellitus (NIDDM)) develops when muscle, fat and liver cells fail to respond normally to insulin. This failure to respond (called insulin resistance) may be due to reduced numbers of insulin 25 receptors on these cells, or a dysfunction of signaling pathways within the cells, or both. The beta cells initially compensate for this insulin resistance by increasing insulin output. Over time, these cells become unable to produce enough insulin to maintain normal glucose levels, indicating progression to Type 2 diabetes.

[0004] Type 2 diabetes is brought on by a combination of genetic and acquired risk factors, 30 including a high-fat diet, lack of exercise, and aging. Worldwide, Type 2 diabetes has become an epidemic, driven by increases in obesity and a sedentary lifestyle, widespread

adoption of western dietary habits, and the general aging of the population in many countries. In 1985, an estimated 30 million people worldwide had diabetes -- by 2000, this figure had increased 5-fold, to an estimated 154 million people. The number of people with diabetes is expected to double between now and 2025, to about 300 million.

5 [0005] Type 2 diabetes is a complex disease characterized by defects in glucose and lipid metabolism. Typically there are perturbations in many metabolic parameters including increases in fasting plasma glucose levels, free fatty acid levels and triglyceride levels, as well as a decrease in the ratio of HDL/LDL. As discussed above, one of the principal underlying causes of diabetes is thought to be an increase in insulin resistance in peripheral 10 tissues, principally muscle and fat.

[0006] Therapies aimed at reducing peripheral insulin resistance are available. The most relevant to this invention are drugs of the thiazolidinedione (TZD) class namely troglitazone, pioglitazone, and rosiglitazone. In the US these have been marketed under the names Rezulin<sup>TM</sup>, Avandia<sup>TM</sup> and Actos<sup>TM</sup>, respectively. The principal effect of these drugs is to 15 improve glucose homeostasis. Notably in diabetics treated with TZDs there are increases in peripheral glucose disposal rates indicative of increased insulin sensitivity in both muscle and fat.

[0007] The molecular target of TZDs is a member of the PPAR family of ligand-activated transcription factors called PPAR gamma. This transcription factor is highly expressed in 20 adipose tissue with much lower levels being observed in muscle. Binding of TZDs to PPAR gamma in target cells and tissues such as fat and muscle brings about a change in gene expression. The link between TZD-altered gene expression in fat and muscle and increased insulin sensitivity is unknown. The present invention addresses this and other problems. The present invention addresses this and other problems.

25

#### BRIEF SUMMARY OF THE INVENTION

[0008] The present invention provides methods for identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods comprise the steps of: 30 (i) contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; and (ii) selecting an agent that modulates the

expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods further comprise selecting an agent that modulates insulin sensitivity.

[0009] In some embodiments, step (ii) comprises selecting an agent that modulates expression of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that modulates the activity of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that specifically binds to the polypeptide. In some embodiments, the polypeptide is expressed in a cell and the cell is contacted with the agent. In some embodiments, the polypeptide is SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. In other embodiments, the polypeptide is SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38.

[0010] The present invention also provides methods of treating a diabetic or pre-diabetic animal. In some embodiments, the methods comprise administering to the animal a therapeutically effective amount of an agent identified as described above. In some embodiments, the agent is an antibody. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the animal is a human.

[0011] The present invention also provides methods of introducing an expression cassette into a cell. In some emodiments, the methods comprise introducing into the cell an expression cassette comprising a promoter operably linked to a polynucleotide encoding a polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0012] In some embodiments, the polypeptide comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. In other embodiments, the polypeptide comprises SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38. In some embodiments, the cell is selected from the group consisting of an adipocyte and a skeletal muscle cell.

[0013] In some embodiments, the methods further comprising introducing the cell into a human. In some embodiments, the human is diabetic. In some embodiments, the human is prediabetic. In some embodiments, the cell is from the human.

[0014] The present invention also provides methods of diagnosing an individual who has 5 Type 2 diabetes or is prediabetic. In some embodiments, the method comprises, detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, 10 SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34; wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is diabetic or prediabetic. In some embodiments, the amino acid sequence comprises SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID 15 NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0015] In some embodiments, the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide.

[0016] In some embodiments, the detecting step comprises quantifying mRNA encoding 20 the polypeptide. In some embodiments, the mRNA is reverse transcribed and amplified in a polymerase chain reaction.

[0017] In some embodiments, the sample is a blood, urine or tissue sample.

[0018] The present invention also provides isolated nucleic acids that hybridize under stringent conditions to a polynucleotide encoding a polypeptide having an amino acid 25 sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34.

[0019] In some embodiments, the nucleic acid is SEQ ID NO:1, SEQ ID NO:5, SEQ ID 30 NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33. In some embodiments, the nucleic acid encodes SEQ ID

NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

[0020] In other embodiments, the nucleic acid is SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:17, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:31, SEQ ID NO:35, or 5 SEQ ID NO:37. In some embodiments, the nucleic acid encodes SEQ ID NO:4, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:36, or SEQ ID NO:38.

[0021] The present invention also provides an expression cassette comprising a heterologous promoter operably linked to a polynucleotide that hybridizes under stringent 10 conditions to a nucleic acid encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34.

[0022] In some embodiments, the polynucleotide comprises SEQ ID NO:1, SEQ ID NO:3, 15 SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37. In some embodiments, the polynucleotide encodes SEQ ID NO:2, SEQ ID NO:4, 20 SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38.

[0023] The present invention also provides host cells transfected with a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an 25 amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34. In some embodiments, the polynucleotide encodes SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, 30 SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38. In some embodiments, the polynucleotide comprises SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ

ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35 or SEQ ID NO:37. In some embodiments, the host cell is a human cell. In other embodiments, the host cell is a bacterium.

5 [0024] The present invention also provides isolated polypeptides comprising an amino acid sequence at least 70% identical to SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:34, SEQ ID NO:10, or SEQ ID NO:28. In some embodiments, the polypeptide is SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 or SEQ ID NO:38.

10

#### DEFINITIONS

[0025] “Insulin sensitivity” refers to the ability of a cell or tissue to respond to insulin. Responses include, e.g., glucose uptake of a cell or tissue in response to insulin stimulation. Sensitivity can be determined at an organismal, tissue or cellular level. For example, blood or urine glucose levels following a glucose tolerance test are indicative of insulin sensitivity. Other methods of measuring insulin sensitivity include, e.g., measuring glucose uptake (see, e.g., Garcia de Herreros, A., and Birnbaum, M. J. *J. Biol. Chem.* 264, 19994-19999 (1989); 20 Klip, A., Li, G., and Logan, W.J. *Am. J. Physiol.* 247, E291-296 (1984)), measuring the glucose infusion rate (GINF) into tissue such as the skeletal muscle (see, e.g., Ludvik *et al.*, *J. Clin. Invest.* 100:2354 (1997); Frias *et al.*, *Diabetes Care* 23:64, (2000)) and measuring sensitivity of GLUT4 translocation (e.g., as described herein) in response to insulin.

[0026] “Activity” of a polypeptide of the invention refers to structural, regulatory, or biochemical functions of a polypeptide in its native cell or tissue. Examples of activity of a polypeptide include both direct activities and indirect activities. Exemplary direct activities are the result of direct interaction with the polypeptide, e.g., enzymatic activity, ligand binding, production or depletion of second messengers (e.g., cAMP, cGMP, IP<sub>3</sub>, DAG, or Ca<sup>2+</sup>), ion flux, phosphorylation levels, transcription levels, and the like. Exemplary indirect activities are observed as a change in phenotype or response in a cell or tissue to a polypeptide’s directed activity, e.g., modulating insulin sensitivity of a cell as a result of the interaction of the polypeptide with other cellular or tissue components.

25

30

[0027] "Predisposition for diabetes" occurs in a person when the person is at high risk for developing diabetes. A number of risk factors are known to those of skill in the art and include: genetic factors (e.g., carrying alleles that result in a higher occurrence of diabetes than in the average population or having parents or siblings with diabetes); overweight (e.g., 5 body mass index (BMI) greater or equal to 25 kg/m<sup>2</sup>); habitual physical inactivity, race/ethnicity (e.g., African-American, Hispanic-American, Native Americans, Asian-Americans, Pacific Islanders); previously identified impaired fasting glucose or impaired glucose tolerance, hypertension (e.g., greater or equal to 140/90 mmHg in adults); HDL cholesterol less than or equal to 35 mg/dl; triglyceride levels greater or equal to 250 mg/dl; a 10 history of gestational diabetes or delivery of a baby over nine pounds; and/or polycystic ovary syndrome. *See, e.g., "Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus" and "Screening for Diabetes" Diabetes Care 25(1): S5-S24 (2002).*

[0028] A "lean individual," when used to compare with a sample from a patient, refers to 15 an adult with a fasting blood glucose level less than 110 mg/dl or a 2 hour PG reading of 140 mg/dl. "Fasting" refers to no caloric intake for at least 8 hours. A "2 hour PG" refers to the level of blood glucose after challenging a patient to a glucose load containing the equivalent of 75g anhydrous glucose dissolved in water. The overall test is generally referred to as an oral glucose tolerance test (OGTT). *See, e.g., Diabetes Care, Supplement 2002, American 20 Diabetes Association: Clinical Practice Recommendations 2002.* The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0029] A "pre-diabetic individual," when used to compare with a sample from a patient, 25 refers to an adult with a fasting blood glucose level greater than 110 mg/dl but less than 126 mg/dl or a 2 hour PG reading of greater than 140 mg/dl but less than 200mg/dl. A "diabetic individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 126 mg/dl or a 2 hour PG reading of greater than 200 mg/dl..

30 [0030] A "diabetes-related nucleic acid" or "diabetes-related polynucleotide" (also referred to as a "nucleic acid of the invention" or a "polynucleotide of the invention") of the invention is a subsequence or full-length polynucleotide sequence of a gene that encodes a polypeptide,

whose activity modulates diabetes or insulin sensitivity, or whose presence or absence is indicative of diabetes or altered insulin sensitivity. Exemplary nucleic acids of the invention include those sequences substantially identical to SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ 5 ID NO:29, or SEQ ID NO:33 or encode polypeptides substantially identical to SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

10 [0031] An "agonist" refers to an agent that binds to, stimulates, increases, activates, facilitates, enhances activation, sensitizes or up regulates the activity or expression of a polypeptide of the invention.

[0032] An "antagonist" refers to an agent that binds to, partially or totally blocks stimulation, decreases, prevents, delays activation, inactivates, desensitizes, or down regulates the activity or expression of a polypeptide of the invention.

15 [0033] "Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the 20 immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

25 [0034] An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) refer to these light and heavy chains respectively.

30 [0035] Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce  $F(ab)'$ <sub>2</sub>, a dimer of Fab which itself is a light chain joined to  $V_H$ - $C_H1$  by a disulfide bond. The  $F(ab)'$ <sub>2</sub> may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the  $F(ab)'$ <sub>2</sub> dimer into an Fab' monomer. The Fab' monomer is essentially

an Fab with part of the hinge region (see, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, 5 the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv).

[0036] The terms "peptidomimetic" and "mimetic" refer to a synthetic chemical compound that has substantially the same structural and functional characteristics of the antagonists or 10 agonists of the invention. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed "peptide mimetics" or "peptidomimetics" (Fauchere, J. *Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p. 392 (1985); and Evans *et al.* *J. Med. Chem.* 30:1229 (1987), which are incorporated herein by reference). Peptide mimetics 15 that are structurally similar to therapeutically useful peptides may be used to produce an equivalent or enhanced therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (i.e., a polypeptide that has a biological or pharmacological activity), such as apolypeptide exemplified in this application, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting 20 of, e.g., -CH<sub>2</sub>NH-, -CH<sub>2</sub>S-, -CH<sub>2</sub>-CH<sub>2</sub>-, -CH=CH- (cis and trans), -COCH<sub>2</sub>-, -CH(OH)CH<sub>2</sub>-, and -CH<sub>2</sub>SO-. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not 25 substantially alter the mimetic's structure and/or activity. For example, a mimetic composition is within the scope of the invention if it is capable of carrying out the binding or other activities of an agonist or antagonist of a polypeptide of the invention.

[0037] The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as 30 well as intervening sequences (introns) between individual coding segments (exons).

[0038] The term "isolated," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is

associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames that flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

5 [0039] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues

10 (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.* (1992); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

15 [0040] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds.

20 [0041] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the

naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline,  $\gamma$ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an  $\alpha$  carbon that is 5 bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical 10 compounds that have a structure that is different from the general chemical structure of an amino acid, but which functions in a manner similar to a naturally occurring amino acid.

[0042] Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

15 [0043] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those nucleic acids that encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of 20 functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid 25 sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in 30 each described sequence.

[0044] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which

alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

5 [0045] The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 10 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 15 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

20 [0046] “Percentage of sequence identity” is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (e.g., a polypeptide of the invention), which does not comprise additions or deletions, for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, 25 dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

30 [0047] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same sequences are substantially identical if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a

comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. The invention provides polypeptides or polynucleotides that are substantially identical to the polypeptides or polynucleotides, respectively, exemplified herein (e.g., SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; or SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33). This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides in length.

[0048] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0049] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

[0050] Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0051] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison

of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0052] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0053] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

[0054] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. The  $T_m$  is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at  $T_m$ , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42°C, or 5X SSC, 1% SDS, incubating at 65°C, with wash in 0.2X SSC, and 0.1% SDS at 55°C, 60°C, or 65°C.

5 Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

[0055] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize 10 under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those 15 of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[0056] The phrase "a nucleic acid sequence encoding" refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a trans-acting regulatory agent. This phrase specifically encompasses degenerate codons (*i.e.*, different 20 codons which encode a single amino acid) of the native sequence or sequences that may be introduced to conform with codon preference in a specific host cell.

[0057] The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic 25 acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[0058] The term "heterologous" when used with reference to portions of a nucleic acid 30 indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to

make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

5 [0059] An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

10 [0060] The phrase "specifically (or selectively) binds to an antibody" or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind 15 in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except 20 for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. *See*, Harlow and Lane *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, NY (1988) for a 25 description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

[0061] "Inhibitors," "activators," and "modulators" of expression or of activity are used to refer to inhibitory, activating, or modulating molecules, respectively, identified using *in vitro* 30 and *in vivo* assays for expression or activity. Modulators encompass *e.g.*, ligands, agonists, antagonists, and their homologs and mimetics. The term "modulator" includes inhibitors and activators. Inhibitors are agents that, *e.g.*, inhibit expression of a polypeptide of the invention

or bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate the activity of a polypeptide of the invention, *e.g.*, antagonists. Activators are agents that, *e.g.*, induce or activate the expression of a polypeptide of the invention or bind to, stimulate, increase, open, activate, facilitate, or 5 enhance activation, sensitize or up regulate the activity of a polypeptide of the invention, *e.g.*, agonists. Modulators include naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, *e.g.*, applying putative modulator compounds to cells expressing a polypeptide of the invention and then determining the functional effects on a polypeptide of the invention activity, as 10 described above. Samples or assays comprising a polypeptide of the invention that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of effect. Control samples (untreated with modulators) are assigned a relative activity value of 100%. Inhibition of a polypeptide of the invention is achieved when the polypeptide activity value 15 relative to the control is about 80%, optionally 50% or 25, 10%, 5% or 1%. Activation of the polypeptide is achieved when the polypeptide activity value relative to the control is 110%, optionally 150%, optionally 200, 300%, 400%, 500%, or 1000-3000% or more higher.

## DETAILED DESCRIPTION OF THE INVENTION

### 20 I. INTRODUCTION

[0062] The present application demonstrates that, surprisingly, modulated levels of mRNA comprising sequences of the invention occur in muscle tissue of insulin-resistant obese, non-diabetic individuals (which population is generally predisposed to become type 2 diabetics) or type 2 diabetic individuals in comparison to the levels in muscle tissue from lean, non- 25 diabetic individuals. Furthermore, in some instances, mRNA levels of sequences described herein in muscle tissue from type 2 diabetic individuals treated with thiazolidinedione (TZD) is changed in comparison to levels of the mRNA in type 2 diabetic individuals before TZD treatment. Therefore, the modulation of the sequences in the study described herein indicates the sequences' involvement in diabetes and pre-diabetes.

30 [0063] Without intending to limit the invention to a particular mechanism of action, it is believed that modulation of the expression or activity of the polypeptides of the invention is beneficial in treating diabetic, pre-diabetic or obese insulin resistant, non-diabetic patients. Furthermore, modulated levels of the polypeptides of the invention are indicative of insulin

resistance. Thus, the detection of a polypeptide of the invention is useful for diagnosis of diabetes and insulin resistance.

[0064] This invention also provides methods of using polypeptides of the invention and modulators of the polypeptides of the invention to diagnose and treat diabetes, pre-diabetes 5 (including insulin resistant individuals) and related metabolic diseases. The present method also provides methods of identifying modulators of expression or activity of the polypeptides of the invention. Such modulators are useful for treating Type 2 diabetes as well as the pathological aspects of diabetes (e.g., insulin resistance).

## 10 II. GENERAL RECOMBINANT NUCLEIC ACID METHODS FOR USE WITH THE INVENTION

[0065] In numerous embodiments of the present invention, nucleic acids encoding a polypeptide of the present invention will be isolated and cloned using recombinant methods. Such embodiments are used, *e.g.*, to isolate polynucleotides identical or substantially identical to SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:15, 15 SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from an polypeptide or polynucleotide of the invention, to monitor gene expression, for the isolation or detection of sequences in different species, for diagnostic purposes in a patient, *e.g.*, to detect mutations in a polypeptide or polynucleotide of the 20 invention or to detect expression levels of nucleic acids or polypeptides. In some embodiments, the sequences encoding the polypeptides of the invention are operably linked to a heterologous promoter. In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, *e.g.*, a human, a mouse, a rat, etc.

### A. General Recombinant Nucleic Acid Methods

25 [0066] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

30 [0067] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa)

or amino acid residue numbers. Protein sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[0068] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & 5 Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et. al.*, *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

[0069] The sequence of the cloned genes and synthetic oligonucleotides can be verified 10 after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace *et al.*, *Gene* 16:21-26 (1981).

#### **B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding Desired Proteins**

[0070] In general, the nucleic acids encoding the subject proteins are cloned from DNA 15 sequence libraries that are made to encode cDNA or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences disclosed herein, which provide a reference for PCR primers and defines suitable regions for isolating probes specific for the polypeptides or polynucleotides of the invention. Alternatively, where the sequence is cloned into an 20 expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against a polypeptide of interest, including those disclosed herein.

[0071] Methods for making and screening genomic and cDNA libraries are well known to those of skill in the art (*see, e.g.*, Gubler and Hoffman *Gene* 25:263-269 (1983); Benton and 25 Davis *Science*, 196:180-182 (1977); and Sambrook, *supra*).

[0072] Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or 30 enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*, and the

recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein *et al.*, *Proc. Natl. Acad. Sci. USA.*, 72:3961-3965 (1975).

[0073] An alternative method combines the use of synthetic oligonucleotide primers with 5 polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific sequences disclosed herein. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be 10 useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a polypeptide of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (see, U.S. Patent Nos. 4,683,195 and 4,683,202). Genes amplified by a PCR reaction can be purified from agarose gels and cloned into an appropriate 15 vector.

[0074] Appropriate primers and probes for identifying the genes encoding a polypeptide of the invention from mammalian tissues can be derived from the sequences provided herein. For a general overview of PCR, see, Innis *et al.* *PCR Protocols: A Guide to Methods and Applications*, Academic Press, San Diego (1990).

[0075] Synthetic oligonucleotides can be used to construct genes. This is done using a 20 series of overlapping oligonucleotides, usually 40-120 bp in length, representing both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

[0076] A polynucleotide encoding a polypeptide of the invention can be cloned using 25 intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes or eukaryotes, using standard methods well known to those of skill in the art.

### III. PURIFICATION OF PROTEINS OF THE INVENTION

[0077] Either naturally occurring or recombinant polypeptides of the invention can be 30 purified for use in functional assays. Naturally occurring polypeptides of the invention can

be purified from any source (e.g., tissues of an organism expressing an ortholog).

Recombinant polypeptides can be purified from any suitable expression system.

[0078] The polypeptides of the invention may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; 5 column chromatography, immunopurification methods, and others (see, e.g., Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al.*, *supra*; and Sambrook *et al.*, *supra*).

[0079] A number of procedures can be employed when recombinant polypeptides are being purified. For example, proteins having established molecular adhesion properties can be 10 reversibly fused to a polypeptide of the invention. With the appropriate ligand, either protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein may be then removed by enzymatic activity. Finally polypeptides can be purified using immunoaffinity columns.

#### A. Purification of Proteins from Recombinant Bacteria

[0080] When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation 15 and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 µg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be ground using a Polytron grinder (Brinkman Instruments, Westbury, NY). Alternatively, the cells can be sonicated on ice. 20 Alternate methods of lysing bacteria are described in Ausubel *et al.* and Sambrook *et al.*, both *supra*, and will be apparent to those of skill in the art.

[0081] The cell suspension is generally centrifuged and the pellet containing the inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, *e.g.*, 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as 25 possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (*e.g.*, 20 mM sodium phosphate, pH 6.8, 150 mM NaCl). Other appropriate buffers will be apparent to those of skill in the art.

[0082] Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer.

5 Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents that are capable of solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins,

10 accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard

15 separation techniques.

[0083] Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (see, Ausubel *et al.*, *supra*). To isolate recombinant proteins from the periplasm, the

20 bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO<sub>4</sub> and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard

25 separation techniques well known to those of skill in the art.

## B. Purification of Proteins from Insect Cells

[0084] Proteins can also be purified from eukaryotic gene expression systems as described in, e.g., Fernandez and Hoeffler, *Gene Expression Systems* (1999). In some embodiments, baculovirus expression systems are used to isolate proteins of the invention. Recombinant

30 baculoviruses are generally generated by replacing the polyhedrin coding sequence of a baculovirus with a gene to be expressed (e.g., encoding a polypeptide of the invention). Viruses lacking the polyhedrin gene have a unique plaque morphology making them easy to

recognize. In some embodiments, a recombinant baculovirus is generated by first cloning a polynucleotide of interest into a transfer vector (e.g., a pUC based vector) such that the polynucleotide is operably linked to a polyhedrin promoter. The transfer vector is transfected with wildtype DNA into an insect cell (e.g., Sf9, Sf21 or BT1-TN-5B1-4 cells), resulting in 5 homologous recombination and replacement of the polyhedrin gene in the wildtype viral DNA with the polynucleotide of interest. Virus can then be generated and plaque purified. Protein expression results upon viral infection of insect cells. Expressed proteins can be harvested from cell supernatant if secreted, or from cell lysates if intracellular. *See, e.g.,* Ausubel *et al.* and Fernandez and Hoeffler, *supra*.

## 10 C. Standard Protein Separation Techniques For Purifying Proteins

### 1. Solubility Fractionation

[0085] Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from 15 the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate 20 to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through either dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol 25 precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

### 2. Size Differential Filtration

[0086] Based on a calculated molecular weight, a protein of greater and lesser size can be 30 isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered

against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

### 3. Column Chromatography

5 [0087] The proteins of interest can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art.

10 [0088] Immunoaffinity chromatography using antibodies raised to a variety of affinity tags such as hemagglutinin (HA), FLAG, Xpress, Myc, hexahistidine (His), glutathione S transferase (GST) and the like can be used to purify polypeptides. The His tag will also act as a chelating agent for certain metals (e.g., Ni) and thus the metals can also be used to purify His-containing polypeptides. After purification, the tag is optionally removed by specific proteolytic cleavage.

15 [0089] It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

## IV. DETECTION OF POLYNUCLEOTIDES OF THE INVENTION

20 [0090] Those of skill in the art will recognize that detection of expression of polynucleotides and polypeptides of the invention has many uses. For example, as discussed herein, detection of levels of polynucleotides and polypeptides of the invention in a patient is useful for diagnosing diabetes or a predisposition for at least some of the pathological effects of diabetes. Moreover, detection of gene expression is useful to identify modulators of expression of polynucleotides and polypeptides of the invention.

25 [0091] A variety of methods of specific DNA and RNA measurement that use nucleic acid hybridization techniques are known to those of skill in the art (*see, Sambrook, supra*). Some methods involve an electrophoretic separation (e.g., Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation (e.g., by dot blot). Southern blot of genomic DNA (e.g., from a human) can be used for screening for restriction fragment length

30

polymorphism (RFLP) to detect the presence of a genetic disorder affecting a polypeptide of the invention.

[0092] The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, 5 common formats include sandwich assays and competition or displacement assays.

Hybridization techniques are generally described in Hames and Higgins *Nucleic Acid Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. U.S.A.*, 63:378-383 (1969); and John *et al.* *Nature*, 223:582-587 (1969).

[0093] Detection of a hybridization complex may require the binding of a signal-generating 10 complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

[0094] The label may also allow indirect detection of the hybridization complex. For 15 example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme molecules to the antibodies or in some cases, by attachment to a radioactive label (see, e.g., Tijssen, "Practice and Theory of Enzyme Immunoassays," *Laboratory Techniques in Biochemistry and Molecular Biology*, Burdon and van Knippenberg Eds., Elsevier (1985), pp. 20 9-20).

[0095] The probes are typically labeled either directly, as with isotopes, chromophores, 25 lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods typically used to detect the presence of hybridized 30 polynucleotides. The most common method of detection is the use of autoradiography with <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P-labeled probes or the like.

[0096] Other labels include, e.g., ligands that bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies that can serve as specific binding pair

members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, NY (1997); and in Haugland *Handbook of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, 5 Inc. (1996).

[0097] In general, a detector that monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a 10 variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

[0098] The amount of, for example, an RNA is measured by quantitating the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of 15 a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation that does not comprise the modulator, or as compared to a baseline established for a particular reaction type. Means of detecting and quantitating labels are well known to those of skill in the art.

[0099] In some embodiments, the target nucleic acid or the probe is immobilized on a solid 20 support. Solid supports suitable for use in the assays of the invention are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement.

[0100] A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPS™), i.e. Gene Chips or 25 microarrays, available from Affymetrix, Inc. in Santa Clara, CA can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. See, Tijssen, *supra.*, Fodor *et al.* (1991) *Science*, 251: 767- 777; Sheldon *et al.* (1993) *Clinical Chemistry* 39(4): 718-719, and Kozal *et al.* (1996) *Nature Medicine* 2(7): 753-759. Similarly, spotted cDNA arrays (arrays of cDNA sequences bound to nylon, glass 30 or another solid support) can also be used to monitor expression of a plurality of genes.

[0101] Typically, the array elements are organized in an ordered fashion so that each element is present at a specified location on the substrate. Because the array elements are at

specified locations on the substrate, the hybridization patterns and intensities (which together create a unique expression profile) can be interpreted in terms of expression levels of particular genes and can be correlated with a particular disease or condition or treatment. See, e.g., Schena *et al.*, *Science* 270: 467-470 (1995)) and (Lockhart *et al.*, *Nature Biotech.*

5 14: 1675-1680 (1996)).

[0102] Hybridization specificity can be evaluated by comparing the hybridization of specificity-control polynucleotide sequences to specificity-control polynucleotide probes that are added to a sample in a known amount. The specificity-control target polynucleotides may have one or more sequence mismatches compared with the corresponding polynucleotide 10 sequences. In this manner, whether only complementary target polynucleotides are hybridizing to the polynucleotide sequences or whether mismatched hybrid duplexes are forming is determined.

[0103] Hybridization reactions can be performed in absolute or differential hybridization formats. In the absolute hybridization format, polynucleotide probes from one sample are 15 hybridized to the sequences in a microarray format and signals detected after hybridization complex formation correlate to polynucleotide probe levels in a sample. In the differential hybridization format, the differential expression of a set of genes in two biological samples is analyzed. For differential hybridization, polynucleotide probes from both biological samples are prepared and labeled with different labeling moieties. A mixture of the two labeled 20 polynucleotide probes is added to a microarray. The microarray is then examined under conditions in which the emissions from the two different labels are individually detectable. Sequences in the microarray that are hybridized to substantially equal numbers of polynucleotide probes derived from both biological samples give a distinct combined 25 fluorescence (Shalon *et al.* PCT publication WO95/35505). In some embodiments, the labels are fluorescent labels with distinguishable emission spectra, such as Cy3 and Cy5 fluorophores.

[0104] After hybridization, the microarray is washed to remove nonhybridized nucleic acids and complex formation between the hybridizable array elements and the polynucleotide probes is detected. Methods for detecting complex formation are well known to those skilled 30 in the art. In some embodiments, the polynucleotide probes are labeled with a fluorescent label and measurement of levels and patterns of fluorescence indicative of complex formation is accomplished by fluorescence microscopy, such as confocal fluorescence microscopy.

[0105] In a differential hybridization experiment, polynucleotide probes from two or more different biological samples are labeled with two or more different fluorescent labels with different emission wavelengths. Fluorescent signals are detected separately with different photomultipliers set to detect specific wavelengths. The relative abundances/expression 5 levels of the polynucleotide probes in two or more samples are obtained.

[0106] Typically, microarray fluorescence intensities can be normalized to take into account variations in hybridization intensities when more than one microarray is used under similar test conditions. In some embodiments, individual polynucleotide probe/target 10 complex hybridization intensities are normalized using the intensities derived from internal normalization controls contained on each microarray.

[0107] Detection of nucleic acids can also be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (e.g., an antibody that is specific for RNA-DNA duplexes). One example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme (typically by recombinant 15 or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. Coutlee *et al.* (1989) *Analytical Biochemistry* 181:153-162; Bogulavski (1986) *et al. J. Immunol. Methods* 89:123-130; Prooijen-Knegt (1982) *Exp. Cell Res.* 141:397-407; Rudkin (1976) *Nature* 265:472-473, Stollar (1970) *PNAS* 65:993-1000; Ballard (1982) *Mol. Immunol.* 19:793-799; Pisetsky and Caster (1982) *Mol.* 20 *Immunol.* 19:645-650; Viscidi *et al.* (1988) *J. Clin. Microbial.* 41:199-209; and Kiney *et al.* (1989) *J. Clin. Microbiol.* 27:6-12 describe antibodies to RNA duplexes, including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, e.g., from Digene Diagnostics, Inc. (Beltsville, MD).

[0108] In addition to available antibodies, one of skill in the art can easily make antibodies 25 specific for nucleic acid duplexes using existing techniques, or modify those antibodies that are commercially or publicly available. In addition to the art referenced above, general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (see, e.g., Paul (ed) *Fundamental Immunology*, *Third Edition* Raven Press, Ltd., NY (1993); Coligan *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and 30 Lane *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites *et al.* (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding *Monoclonal Antibodies: Principles and Practice* (2d ed.)

Academic Press, New York, NY, (1986); and Kohler and Milstein *Nature* 256: 495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (see, Huse *et al.* *Science* 246:1275-1281 (1989); and Ward *et al.* *Nature* 341:544-546 (1989)). Specific monoclonal and polyclonal antibodies and antisera will usually bind with a  $K_D$  of at least about 0.1  $\mu\text{M}$ , preferably at least about 0.01  $\mu\text{M}$  or better, and most typically and preferably, 0.001  $\mu\text{M}$  or better.

5 [0109] The nucleic acids used in this invention can be either positive or negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

10 [0110] The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario) and Q Beta 15 Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, 20 nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation. It is understood that various detection probes, including Taqman and molecular beacon probes can be used to monitor amplification reaction products, e.g., in real time.

25 [0111] An alternative means for determining the level of expression of the nucleic acids of the present invention is *in situ* hybridization. *In situ* hybridization assays are well known and are generally described in Angerer *et al.*, *Methods Enzymol.* 152:649-660 (1987). In an *in situ* hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the 30 cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

[0112] Single nucleotide polymorphism (SNP) analysis is also useful for detecting differences between alleles of the polynucleotides (e.g., genes) of the invention. SNPs linked to genes encoding polypeptides of the invention are useful, for instance, for diagnosis of diseases (e.g., diabetes) whose occurrence is linked to the gene sequences of the invention.

5 For example, if an individual carries at least one SNP linked to a disease-associated allele of the gene sequences of the invention, the individual is likely predisposed for one or more of those diseases. If the individual is homozygous for a disease-linked SNP, the individual is particularly predisposed for occurrence of that disease (e.g., diabetes). In some 10 embodiments, the SNP associated with the gene sequences of the invention is located within 300,000; 200,000; 100,000; 75,000; 50,000; or 10,000 base pairs from the gene sequence.

[0113] Various real-time PCR methods including, e.g., Taqman or molecular beacon-based assays (e.g., U.S. Patent Nos. 5,210,015; 5,487,972; Tyagi *et al.*, *Nature Biotechnology* 14:303 (1996); and PCT WO 95/13399 are useful to monitor for the presence of absence of a SNP. Additional SNP detection methods include, e.g., DNA sequencing, sequencing by 15 hybridization, dot blotting, oligonucleotide array (DNA Chip) hybridization analysis, or are described in, e.g., U.S. Patent No. 6,177,249; Landegren *et al.*, *Genome Research*, 8:769-776 (1998); Botstein *et al.*, *Am J Human Genetics* 32:314-331 (1980); Meyers *et al.*, *Methods in Enzymology* 155:501-527 (1987); Keen *et al.*, *Trends in Genetics* 7:5 (1991); Myers *et al.*, *Science* 230:1242-1246 (1985); and Kwok *et al.*, *Genomics* 23:138-144 (1994).

20 **V. IMMUNOLOGICAL DETECTION OF POLYPEPTIDES OF THE INVENTION**

[0114] In addition to the detection of polynucleotides of the invention and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect 25 polypeptides of the invention. Immunoassays can be used to qualitatively or quantitatively analyze polypeptides of the invention. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

**A. Antibodies to Target Proteins or other immunogens**

[0115] Methods for producing polyclonal and monoclonal antibodies that react specifically with a protein of interest or other immunogen are known to those of skill in the art (see, e.g., 30 Coligan, *supra*; and Harlow and Lane, *supra*; Stites *et al.*, *supra* and references cited therein; Goding, *supra*; and Kohler and Milstein *Nature*, 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant

antibodies in phage or similar vectors (see, Huse *et al.*, *supra*; and Ward *et al.*, *supra*). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is 5 immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[0116] Polyclonal sera are collected and titered against the immunogen in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid 10 support. Polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their crossreactivity against proteins other than the polypeptides of the invention or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a  $K_D$  of at least about 0.1 mM, more usually at least about 1  $\mu$ M, preferably at least about 0.1  $\mu$ M or 15 better, and most preferably, 0.01  $\mu$ M or better.

[0117] A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure 20 form. Synthetic peptides made using the protein sequences described herein may also be used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described *supra*. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays 25 to measure the protein.

[0118] Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to polypeptides of the 30 invention. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to

enrich for antibodies reactive to the protein can be done if desired (see, Harlow and Lane, *supra*).

[0119] Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen 5 are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, e.g., transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the 10 monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences that encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *supra*.

[0120] Once target immunogen-specific antibodies are available, the immunogen can be measured by a variety of immunoassay methods with qualitative and quantitative results available to the clinician. For a review of immunological and immunoassay procedures in general see, Stites, *supra*. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio 15 *Enzyme Immunoassay*, CRC Press, Boca Raton, Florida (1980); Tijssen, *supra*; and Harlow and Lane, *supra*.

[0121] Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum that was raised to full-length polypeptides of the invention or a fragment thereof. This antiserum is selected to have low cross-reactivity against other proteins and any such 25 cross-reactivity is removed by immunoabsorption prior to use in the immunoassay.

## B. Immunological Binding Assays

[0122] In some embodiments, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (see, e.g., U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also 30 Asai *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Academic Press, Inc. NY (1993); Stites, *supra*. Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (e.g., full-length

polypeptides of the present invention, or antigenic subsequences thereof). The capture agent is a moiety that specifically binds to the analyte. The antibody may be produced by any of a number of means well known to those of skill in the art and as described above.

[0123] Immunoassays also often utilize a labeling agent to bind specifically to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

[0124] In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

[0125] Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (see, generally, Kronval, *et al.* *J. Immunol.*, 111:1401-1406 (1973); and Akerstrom, *et al.* *J. Immunol.*, 135:2589-2542 (1985)).

[0126] Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

#### 1. Non-Competitive Assay Formats

[0127] Immunoassays for detecting proteins or analytes of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured protein or analyte is directly measured. In one preferred "sandwich" assay, for example, the capture agent (*e.g.*, antibodies specific for the polypeptides of the invention) can be bound directly to a solid substrate where it is immobilized. These immobilized antibodies then capture the polypeptide present in the test sample. The

polypeptide of the invention thus immobilized is then bound by a labeling agent, such as a second labelled antibody specific for the polypeptide. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be 5 modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

## 2. Competitive Assay Formats

[0128] In competitive assays, the amount of protein or analyte present in the sample is measured indirectly by measuring the amount of an added (exogenous) protein or analyte 10 displaced (or competed away) from a specific capture agent (e.g., antibodies specific for a polypeptide of the invention) by the protein or analyte present in the sample. The amount of immunogen bound to the antibody is inversely proportional to the concentration of immunogen present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of analyte may be detected by providing a 15 labeled analyte molecule. It is understood that labels can include, e.g., radioactive labels as well as peptide or other tags that can be recognized by detection reagents such as antibodies.

[0129] Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay and compete with the 20 binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above are selected and pooled. The 25 cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, e.g., distantly related homologs.

[0130] The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps a protein of the present invention, to the immunogen protein. In order to make this 30 comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than

10 times the amount of the protein partially encoded by a sequence herein that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

### 3. Other Assay Formats

5 [0131] In some embodiments, western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide of the invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, *e.g.*, a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample  
10 with the antibodies that specifically bind the protein of interest. For example, antibodies are selected that specifically bind to the polypeptides of the invention on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

15 [0132] Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see*, Monroe *et al.* (1986) *Amer. Clin. Prod. Rev.* 5:34-41).

### 4. Labels

20 [0133] The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the  
25 present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, Dynabeads<sup>TM</sup>), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others  
30 commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, *etc.*) beads.

[0134] The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

[0135] Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, *see, e.g.*, U.S. Patent No. 4,391,904).

[0136] Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product.

20 Finally simple colorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

[0137] Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, 25 antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

## VI. IDENTIFICATION OF MODULATORS OF POLYPEPTIDES OF THE INVENTION

30 [0138] Modulators of a polypeptide of the invention, *i.e.* agonists or antagonists of a polypeptide's activity, or polypeptide's or polynucleotide's expression, are useful for treating a number of human diseases, including diabetes. For example, administration of modulators

can be used to treat diabetic patients or prediabetic individuals to prevent progression, and therefore symptoms, associated with diabetes (including insulin resistance).

#### A. Agents that Modulate Polypeptides of the Invention

[0139] The agents tested as modulators of polypeptides of the invention can be any small

5 chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid.

Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large

10 chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). Modulators also include agents designed to reduce the level of mRNA encoding a polypeptide of the invention (e.g. antisense molecules, ribozymes, DNAzymes, small inhibitory RNAs and the like) or the level of translation from an mRNA

15 (e.g., translation blockers such as an antisense molecules that are complementary to translation start or other sequences on an mRNA molecule). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

20 [0140] In some embodiments, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired 25 characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[0141] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical 30 library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of

amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[0142] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, 5 peptide libraries (see, e.g., U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), 10 benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen 15 *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (see Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid 20 libraries (see, e.g., U.S. Patent 5,539,083), antibody libraries (see, e.g., Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 25 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like). 30 [0143] Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, etc.).

**B. Methods of Screening for Modulators of the Polypeptides of the Invention**

[0144] A number of different screening protocols can be utilized to identify agents that modulate the level of expression or activity of a polynucleotide of a polypeptide of the invention in cells, particularly mammalian cells, and especially human cells. In general 5 terms, the screening methods involve screening a plurality of agents to identify an agent that modulates the activity of a polypeptide of the invention by, e.g., binding to the polypeptide, preventing an inhibitor or activator from binding to the polypeptide, increasing association of an inhibitor or activator with the polypeptide, or activating or inhibiting expression of the polypeptide.

10 [0145] Any cell expressing a full-length polypeptide of the invention or a fragment thereof can be used to identify modulators. In some embodiments, the cells are eukaryotic cells lines (e.g., CHO or HEK293) transformed to express a heterologous polypeptide of the invention. In some embodiments, a cell expressing an endogenous polypeptide of the invention is used in screens. In other embodiments, modulators are screened for their ability to effect insulin 15 responses.

1. Polypeptide Binding Assays

[0146] Preliminary screens can be conducted by screening for agents capable of binding to polypeptides of the invention, as at least some of the agents so identified are likely modulators of a polypeptide of the invention. Binding assays are also useful, e.g., for 20 identifying endogenous proteins that interact with polypeptides of the invention. For example, antibodies, receptors or other molecules that bind polypeptides of the invention can be identified in binding assays.

[0147] Binding assays usually involve contacting a polypeptide of the invention with one or more test agents and allowing sufficient time for the protein and test agents to form a 25 binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, methods that measure co-precipitation or co-migration on non-denaturing SDS-polyacrylamide gels, and co-migration on Western blots (see, e.g., Bennet, J.P. and Yamamura, H.I. (1985) "Neurotransmitter, Hormone or Drug Receptor Binding Methods," in 30 *Neurotransmitter Receptor Binding* (Yamamura, H. I., *et al.*, eds.), pp. 61-89. Other binding assays involve the use of mass spectrometry or NMR techniques to identify molecules bound

to a polypeptide of the invention or displacement of labeled substrates. The polypeptides of the invention utilized in such assays can be naturally expressed, cloned or synthesized.

[0148] In addition, mammalian or yeast two-hybrid approaches (*see, e.g.*, Bartel, P.L. *et. al. Methods Enzymol*, 254:241 (1995)) can be used to identify polypeptides or other 5 molecules that interact or bind when expressed together in a host cell.

## 2. Polypeptide Activity

[0149] The activity of polypeptides of the invention can be assessed using a variety of *in vitro* and *in vivo* assays to determine functional, chemical, and physical effects, *e.g.*, measuring ligand binding (*e.g.*, radioactive or otherwise labeled ligand binding), second 10 messengers (*e.g.*, cAMP, cGMP, IP<sub>3</sub>, DAG, or Ca<sup>2+</sup>), ion flux, phosphorylation levels, transcription levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of the polypeptides of the invention. Modulators can also be genetically altered versions of polypeptides of the invention.

[0150] The polypeptide of the assay will be selected from a polypeptide with substantial 15 identity to a sequence described herein, *e.g.*, SEQ ID NO:2, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:16, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, or other conservatively modified variants thereof. Generally, the amino acid sequence identity will be at least 70%, optionally at least 85%, optionally at least 90-95% to the polypeptides exemplified herein. Optionally, the polypeptide of the assays 20 will comprise a fragment of a polypeptide of the invention, such as an extracellular domain, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. Either a polypeptide of the invention or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

[0151] Modulators of polypeptide activity are tested using either recombinant or naturally 25 occurring polypeptides of the invention. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, tissue slices, dissociated cells, *e.g.*, from tissues expressing polypeptides of the invention, transformed cells, or membranes can be used. Modulation is tested using one of the *in vitro* or *in vivo* assays described herein.

[0152] Modulator binding to polypeptides of the invention, a domain, or chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid

monolayer, or in vesicles. Binding of a modulator can be tested using, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbance, refractive index), hydrodynamic (*e.g.*, shape), chromatographic, or solubility properties.

[0153] Samples or assays that are treated with a potential modulator (*e.g.*, a “test compound”) are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative activity value of 100. Inhibition of the polypeptides of the invention is achieved when the activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of the polypeptides of the invention is achieved when the activity value relative to the control is 110%, optionally 150%, 200%, 300%, 400%, 500%, or 1000-2000%.

### 3. Expression Assays

[0154] Screening for a compound that modulates the expression of a polynucleotide or a polypeptide of the invention is also provided. Screening methods generally involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing a polynucleotide or a polypeptide of the invention, and then detecting an increase or decrease in expression (either transcript or translation product). Assays can be performed with any cells that express a polynucleotide or a polypeptide of the invention.

[0155] Expression can be detected in a number of different ways. As described *infra*, the expression level of a polynucleotide of the invention in a cell can be determined by probing the mRNA expressed in a cell with a probe that specifically hybridizes with a transcript (or complementary nucleic acid derived therefrom) of a polynucleotide of the invention. Probing can be conducted by lysing the cells and conducting Northern blots or without lysing the cells using *in situ*-hybridization techniques. Alternatively, a polypeptide of the invention can be detected using immunological methods in which a cell lysate is probed with antibodies that specifically bind to the polypeptide.

[0156] The level of expression or activity of a polynucleotide or a polypeptide of the invention can be compared to a baseline value. The baseline value can be a value for a control sample or a statistical value that is representative of expression levels of a polynucleotide or a polypeptide of the invention for a control population (*e.g.*, lean individuals as described herein) or cells (*e.g.*, tissue culture cells not exposed to a modulator). Expression levels can also be determined for cells that do not express the polynucleotide or a

polypeptide of the invention as a negative control. Such cells generally are otherwise substantially genetically the same as the test cells.

[0157] A variety of different types of cells can be utilized in the reporter assays. Cells that do not endogenously express a polypeptide of the invention can be prokaryotic, but are 5 preferably eukaryotic. The eukaryotic cells can be any of the cells typically utilized in generating cells that harbor recombinant nucleic acid constructs. Exemplary eukaryotic cells include, but are not limited to, yeast, and various higher eukaryotic cells such as the HEK293, HepG2, COS, CHO and HeLa cell lines.

[0158] Various controls can be conducted to ensure that an observed activity is authentic 10 including running parallel reactions with cells that lack the reporter construct or by not contacting a cell harboring the reporter construct with test compound. Compounds can also be further validated as described below.

#### 4. Validation

[0159] Agents that are initially identified by any of the foregoing screening methods can be 15 further tested to validate the apparent activity. Modulators that are selected for further study can be tested on the "classic" insulin responsive cell line, mouse 3T3-L1 adipocytes, muscle cells such as L6 cells and the like. Cells (e.g., adipocytes or muscle cells) are pre-incubated with the modulators and tested for acute (up to 4 hours) and chronic (overnight) effects on basal and insulin-stimulated GLUT4 translocation and glucose uptake.

[0160] Following such studies, validity of the modulators is tested in suitable animal 20 models. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model for humans and then determining if expression of activity of a polypeptide of the invention is in fact modulated.

[0161] The effect of the compound will be assessed in either diabetic animals or in diet 25 induced insulin resistant animals. The blood glucose and insulin levels will be determined. The animal models utilized in validation studies generally are mammals of any kind. Specific examples of suitable animals include, but are not limited to, primates, mice and rats. For example, monogenic models of diabetes (e.g., ob/ob and db/db mice, Zucker rats and Zucker Diabetic Fatty rats etc) or polygenic models of diabetes (e.g., OLETF rats, GK rats, 30 NSY mice, and KK mice) can be useful for validating modulation of a polypeptide of the invention in a diabetic or insulin resistant animal. In addition, transgenic animals expressing human polypeptides of the invention can be used to further validate drug candidates.

### C. Solid Phase and Soluble High Throughput Assays

[0162] In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, 5 if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 or more different compounds are possible using 10 the integrated systems of the invention. In addition, microfluidic approaches to reagent manipulation can be used.

[0163] A molecule of interest (e.g., a polypeptide or polynucleotide of the invention, or a modulator thereof) can be bound to the solid-state component, directly or indirectly, via covalent or non-covalent linkage, e.g., via a tag. The tag can be any of a variety of 15 components. In general, a molecule that binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

[0164] A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, 20 for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, poly-His, etc.) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (see, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

[0165] Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody that recognizes the first antibody. In addition to antibody-antigen 30 interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors,

interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; *see, e.g.*, Pigott & Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, *etc.*), intracellular receptors (*e.g.*, which mediate the effects of various small 5 ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

[0166] Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, 10 polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

[0167] Common linkers such as peptides, polyethers, and the like can also serve as tags, 15 and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc., Huntsville, Alabama. These linkers optionally have amide linkages, sulphydryl linkages, or heterofunctional linkages.

[0168] Tag binders are fixed to solid substrates using any of a variety of methods currently 20 available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent that fixes a chemical group to the surface that is reactive with a portion of the tag binder. For example, groups that are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups.

25 Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature (*see, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, *e.g.*, peptides); Geysen *et al.*, *J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, 30 *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al.*, *Science*, 251:767-777 (1991); Sheldon *et al.*, *Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al.*, *Nature Medicine* 2(7):753759 (1996) (all describing

arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

[0169] The invention provides *in vitro* assays for identifying, in a high throughput format, 5 compounds that can modulate the expression or activity of a polypeptide of the invention.

Control reactions that measure activity of a polypeptide of the invention in a cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform.

Such optional control reactions are appropriate and increase the reliability of the assay.

Accordingly, in some embodiments, the methods of the invention include such a control

10 reaction. For each of the assay formats described, “no modulator” control reactions that do not include a modulator provide a background level of binding activity.

[0170] In some assays it will be desirable to have positive controls. At least two types of 15 positive controls are appropriate. First, a known activator of a polypeptide or a polynucleotide of the invention can be incubated with one sample of the assay, and the

resulting increase in signal resulting from an increased expression level or activity of a 20 polypeptide or a polynucleotide of the invention are determined according to the methods herein. Second, a known inhibitor of a polypeptide or a polynucleotide of the invention can be added, and the resulting decrease in signal for the expression or activity of a polypeptide or a polynucleotide of the invention can be similarly detected. It will be appreciated that

modulators can also be combined with activators or inhibitors to find modulators that inhibit 25 the increase or decrease that is otherwise caused by the presence of the known modulator of a polypeptide or a polynucleotide of the invention.

## VII. COMPOSITIONS, KITS AND INTEGRATED SYSTEMS

[0171] The invention provides compositions, kits and integrated systems for practicing the 25 assays described herein using nucleic acids or polypeptides of the invention, antibodies, etc.

[0172] The invention provides assay compositions for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a polypeptide of the invention immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization.

30 Modulators of expression or activity of a polypeptide of the invention can also be included in the assay compositions.

[0173] The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises an antibody that specifically binds to a polypeptide of the invention or a polynucleotide sequence encoding such polypeptides, and a label for detecting the presence of the probe. The kits may include at least one 5 polynucleotide sequence encoding a polypeptide of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding a polypeptide of the invention, or on activity of a polypeptide of the invention, one or more containers or compartments (e.g., to hold the probe, labels, or the 10 like), a control modulator of the expression or activity of a polypeptide of the invention, a robotic armature for mixing kit components or the like.

[0174] The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of a polypeptide of the invention. The systems can include a robotic armature which transfers fluid from a source to 15 a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

[0175] A number of robotic fluid transfer systems are available, or can easily be made from 20 existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, MA) automated robot using a Microlab 2200 (Hamilton; Reno, NV) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous binding assays.

[0176] Optical images viewed (and, optionally, recorded) by a camera or other recording 25 device (e.g., a photodiode and data storage device) are optionally further processed in any of the embodiments herein, e.g., by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image.

[0177] One conventional system carries light from the specimen field to a cooled charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array 30 of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (e.g., individual hybridization sites on an

array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, *e.g.*, by fluorescent or dark field microscopic techniques.

## 5      **VIII. ADMINISTRATION AND PHARMACEUTICAL COMPOSITIONS**

[0178] Modulators of the polypeptides of the invention (*e.g.*, antagonists or agonists) can be administered directly to the mammalian subject for modulation of activity of a polypeptide of the invention *in vivo*. Administration is by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated and is well known to 10 those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0179] The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the 15 particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g.*, *Remington's Pharmaceutical Sciences*, 17<sup>th</sup> ed. 1985)).

[0180] The modulators (*e.g.*, agonists or antagonists) of the expression or activity of the a 20 polypeptide of the invention, alone or in combination with other suitable components, can be prepared for injection or for use in a pump device. Pump devices (also known as "insulin pumps") are commonly used to administer insulin to patients and therefore can be easily adapted to include compositions of the present invention. Manufacturers of insulin pumps include Animas, Disetronic and MiniMed.

[0181] The modulators (*e.g.*, agonists or antagonists) of the expression or activity of a 25 polypeptide of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0182] Formulations suitable for administration include aqueous and non-aqueous 30 solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and

solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or intrathecally. The 5 formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part of a prepared food or drug.

[0183] The dose administered to a patient, in the context of the present invention should be 10 sufficient to induce a beneficial response in the subject over time. The optimal dose level for any patient will depend on a variety of factors including the efficacy of the specific modulator employed, the age, body weight, physical activity, and diet of the patient, on a possible combination with other drugs, and on the severity of the case of diabetes. It is recommended that the daily dosage of the modulator be determined for each individual 15 patient by those skilled in the art in a similar way as for known insulin compositions. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

[0184] In determining the effective amount of the modulator to be administered a physician 20 may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

[0185] For administration, modulators of the present invention can be administered at a rate 25 determined by the LD-50 of the modulator, and the side-effects of the modulator at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

[0186] The compounds of the present invention can also be used effectively in combination with one or more additional active agents depending on the desired target therapy (see, e.g., Turner, N. et al. *Prog. Drug Res.* (1998) 51: 33-94; Haffner, S. *Diabetes Care* (1998) 21: 30 160-178; and DeFronzo, R. et al. (eds.), *Diabetes Reviews* (1997) Vol. 5 No. 4). A number of studies have investigated the benefits of combination therapies with oral agents (see, e.g., Mahler, R., *J. Clin. Endocrinol. Metab.* (1999) 84: 1165-71; United Kingdom Prospective

Diabetes Study Group: UKPDS 28, *Diabetes Care* (1998) 21: 87-92; Bardin, C. W.,(ed.), *Current Therapy In Endocrinology And Metabolism*, 6th Edition (Mosby - Year Book, Inc., St. Louis, MO 1997); Chiasson, J. et al., *Ann. Intern. Med.* (1994) 121: 928-935; Coniff, R. et al., *Clin. Ther.* (1997) 19: 16-26; Coniff, R. et al., *Am. J. Med.* (1995) 98: 443-451; and

5 Iwamoto, Y. et al., *Diabet. Med.* (1996) 13 365-370; Kwiterovich, P. *Am. J. Cardiol* (1998) 82(12A): 3U-17U). These studies indicate that modulation of diabetes, among other diseases, can be further improved by the addition of a second agent to the therapeutic regimen.

Combination therapy includes administration of a single pharmaceutical dosage formulation that contains a modulator of the invention and one or more additional active agents, as well as

10 administration of a modulator and each active agent in its own separate pharmaceutical dosage formulation. For example, a modulator and a thiazolidinedione can be administered to the human subject together in a single oral dosage composition, such as a tablet or capsule, or each agent can be administered in separate oral dosage formulations. Where separate dosage formulations are used, a modulator and one or more additional active agents can be

15 administered at essentially the same time (i.e., concurrently), or at separately staggered times (i.e., sequentially). Combination therapy is understood to include all these regimens.

[0187] One example of combination therapy can be seen in treating pre-diabetic individuals (e.g., to prevent progression into type 2 diabetes) or diabetic individuals (or treating diabetes and its related symptoms, complications, and disorders), wherein the modulators can be

20 effectively used in combination with, for example, sulfonylureas (such as chlorpropamide, tolbutamide, acetohexamide, tolazamide, glyburide, gliclazide, glynase, glimepiride, and glipizide); biguanides (such as metformin); a PPAR beta delta agonist; a ligand or agonist of PPAR gamma such as thiazolidinediones (such as ciglitazone, pioglitazone (see, e.g., U.S. Patent No. 6,218,409), troglitazone, and rosiglitazone (see, e.g., U.S. Patent No. 5,859,037));

25 PPAR alpha agonists such as clofibrate, gemfibrozil, fenofibrate, ciprofibrate, and bezafibrate; dehydroepiandrosterone (also referred to as DHEA or its conjugated sulphate ester, DHEA-SO<sub>4</sub>); antiglucocorticoids; TNF $\alpha$  inhibitors;  $\alpha$ -glucosidase inhibitors (such as acarbose, miglitol, and voglibose); amylin and amylin derivatives (such as pramlintide, (see, also, U.S. Patent Nos. 5,902,726; 5,124,314; 5,175,145 and 6,143,718.)); insulin

30 secretagogues (such as repaglinide, gliclizide, and nateglinide (see, also, U.S. Patent Nos. 6,251,856; 6,251,865; 6,221,633; 6,174,856)), and insulin.

## IX. GENE THERAPY

[0188] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered polypeptides of the invention in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding 5 polypeptides of the invention to cells *in vitro*. In some embodiments, the nucleic acids encoding polypeptides of the invention are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes 10 after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative 15 Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

[0189] Methods of non-viral delivery of nucleic acids encoding engineered polypeptides of the invention include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, 20 and agent-enhanced uptake of DNA. Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* 25 administration).

[0190] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); 30 Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0191] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered polypeptides of the invention take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be 5 used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of polypeptides of the invention could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene 10 transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies 15 have been observed in many different cell types and target tissues.

[0192] The tropism of a retrovirus can be altered by incorporating foreign envelope 20 proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with 25 packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape 30 leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al.*, *J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*, *Virol.* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0193] In applications where transient expression of the polypeptides of the invention is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable 35 of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures

(see, e.g., West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994)). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); 5 Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989).

[0194] pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al.*, *Blood* 85:3048-305 (1995); Kohn *et al.*, *Nat. Med.* 1:1017-102 (1995); Malech *et al.*, *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first 10 therapeutic vector used in a gene therapy trial. (Blaese *et al.*, *Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al.*, *Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al.*, *Hum. Gene Ther.* 1:111-2 (1997)).

[0195] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative 15 gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al.*, *Lancet* 351:9117 1702-3 (1998), Kearns *et* 20 *al.*, *Gene Ther.* 9:748-55 (1996)).

[0196] Replication-deficient recombinant adenoviral vectors (Ad) can be engineered such 25 that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues *in vivo*, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al.*, *Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include 30 Rosenecker *et al.*, *Infection* 24:1 5-10 (1996); Sterman *et al.*, *Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al.*, *Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al.*, *Hum. Gene*

*Ther.* 5:597-613 (1997); Topf *et al.*, *Gene Ther.* 5:507-513 (1998); Sterman *et al.*, *Hum. Gene Ther.* 7:1083-1089 (1998).

[0197] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and  $\psi$ 2 cells or PA317 cells, 5 which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging 10 cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the 15 AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0198] In many gene therapy applications, it is desirable that the gene therapy vector be 20 delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be 25 modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular 30 receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

[0199] Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient 5 (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0200] *Ex vivo* cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the 10 art. In some embodiments, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA) encoding a polypeptides of the invention, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (see, e.g., Freshney *et al.*, *Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion 15 of how to isolate and culture cells from patients).

[0201] In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* 20 into clinically important immune cell types using cytokines such a GM-CSF, IFN- $\gamma$  and TNF- $\alpha$  are known (see Inaba *et al.*, *J. Exp. Med.* 176:1693-1702 (1992)).

[0202] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow 25 cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (see Inaba *et al.*, *J. Exp. Med.* 176:1693-1702 (1992)).

[0203] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes 30 normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular

composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0204] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989).

## X. DIAGNOSIS OF DIABETES

[0205] The present invention also provides methods of diagnosing diabetes or a predisposition of at least some of the pathologies of diabetes. Diagnosis can involve determination of a genotype of an individual (e.g., with SNPs) and comparison of the genotype with alleles known to have an association with the occurrence of diabetes. Alternatively, diagnosis also involves determining the level of a polypeptide or polynucleotide of the invention in a patient and then comparing the level to a baseline or range. Typically, the baseline value is representative of a polypeptide or polynucleotide of the invention in a healthy (e.g., lean) person.

[0206] As discussed above, variation of levels (e.g., low or high levels) of a polypeptide or polynucleotide of the invention compared to the baseline range indicates that the patient is either diabetic or at risk of developing at least some of the pathologies of diabetes (e.g., pre-diabetic). The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0207] In some embodiments, the level of polypeptide or polynucleotide of the invention is measured by taking a blood, urine or tissue sample from a patient and measuring the amount of a polypeptide or polynucleotide of the invention in the sample using any number of detection methods, such as those discussed herein. For instance, fasting and fed blood or urine levels can be tested.

[0208] In some embodiments, the baseline level and the level in a lean sample from an individual, or at least two samples from the same individual differ by at least about 5%, 10%, 20%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500%, 1000% or more. In some

embodiments, the sample from the individual is greater by at least one of the above-listed percentages relative to the baseline level. In some embodiments, the sample from the individual is lower by at least one of the above-listed percentages relative to the baseline level.

5 [0209] In some embodiments, the level of a polypeptide or polynucleotide of the invention is used to monitor the effectiveness of antidiabetic therapies such as thiazolidinediones, metformin, sulfonylureas and other standard therapies. In some embodiments the activity or expression of a polypeptide or polynucleotide of the invention will be measured prior to and after treatment of diabetic or pre-diabetic patients with antidiabetic therapies as a surrogate 10 marker of clinical effectiveness. For example, the greater the reduction in expression or activity of a polypeptide of the invention indicates greater effectiveness.

[0210] Glucose/insulin tolerance tests can also be used to detect the effect of glucose levels on levels of a polypeptide or polynucleotide of the invention. In glucose tolerance tests, the patient's ability to tolerate a standard oral glucose load is evaluated by assessing serum and 15 urine specimens for glucose levels. Blood samples are taken before the glucose is ingested, glucose is given by mouth, and blood or urine glucose levels are tested at set intervals after glucose ingestion. Similarly, meal tolerance tests can also be used to detect the effect of insulin or food, respectively, on levels of a polypeptide or polynucleotide of the invention.

20 [0211] All publications, accession numbers, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0212] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to 25 one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

#### EXAMPLES

[0213] The following examples are offered to illustrate, but not to limit the claimed 30 invention.

[0214] In obese insulin-resistant or type II diabetic individuals, peripheral tissues especially muscle and fat are known to have a moderately impaired ability to respond to insulin and hence to take up glucose. This defect in glucose metabolism is usually compensated for by increased secretion of insulin from the pancreas, thereby maintaining normal glucose levels.

5 The majority of glucose disposal occurs in the muscle. A number of obese insulin-resistant patients will progress to overt diabetics over time. The molecular defects underlying this peripheral insulin resistance in obese individuals are not well defined. Genes in muscle or fat that exhibit altered expression in obese individuals when compared to lean individuals can be causative genes for insulin resistance and are also able to predict the transition to diabetes.

10 Modulators of such genes can reduce or reverse insulin resistance and increase or restore insulin sensitivity to normal, thereby improving whole body glucose homeostasis, including; for example; insulin secretion. Modulators of such genes can also be used to pre-empt the transition from obesity-induced insulin resistance to diabetes. For these reasons, gene expression profiling was performed in muscle from lean and obese individuals.

15 [0215] The molecular mechanism by which thiazolidinediones (TZDs) cause an increase in peripheral insulin sensitivity was studied. Genes in muscle or fat whose expression is altered by TZDs may lie on a pathway leading from TZD treatment to increased insulin sensitivity. Modulators of such genes can elicit the same effect as TZD treatment. Moreover, such modulators can lack some of the side effects of TZD. The majority of glucose disposal

20 occurs in muscle. For this reason, gene expression profiling in human muscle from diabetics treated with troglitazone was used to identify genes important for TZD action and therefore treatment of diabetes and insulin resistance.

[0216] Gene expression profiling was performed on tissue samples (muscle) obtained from lean obese and diabetic individuals. Two studies were performed. In the first study, basal

25 samples were isolated from all individuals at the beginning of a 5 hour hyperinsulinemic euglycemic clamp. Clamp samples were isolated at the end of this procedure. Similar basal and clamp samples were taken 3 months later after all patients had taken the insulin sensitizing drug troglitazone (tro).

[0217] In the second study, samples were obtained from lean, obese, and diabetic

30 individuals before and after a hyperinsulinemic euglycemic clamp. No troglitazone treatment was used. For all tissue samples mRNA was isolated from these muscle samples and

converted to cRNA by standard procedures. The gene expression profile for each individual was determined by hybridization of cRNA to custom synthesized Affymetrix chips.

[0218] Gene expression profile differences were calculated as follows. The expression level of a particular gene is indicated by its 'average difference score'. The raw data was analyzed by a statistical test to remove 'outliers'. The mean 'average difference score' was then calculated from the average difference scores for all individuals in a particular treatment group. Genes were determined to be changed in three different comparisons by calculating the Students t test statistic between two conditions and selecting those with t less than or equal to 0.05. Fold change was determined as the ratio of mean average difference score in condition 2 to the mean average difference score in condition 1. The first comparison is basal leans (condition 1) versus basal obese (condition 2). The second comparison is basal leans (condition 1) versus basal diabetics (condition2). The third comparison is basal diabetics before troglitazone treatment (condition 1) versus basal diabetics after troglitazone treatment (condition 2).

[0219] The gene expression changes observed in the human muscle samples derived from the various comparison populations analyzed in the profiling analysis described above were further confirmed by the use of real-time quantitative PCR. A combination of specific PCR primers and a Taqman probe were designed to detect and quantify expression levels for each gene. Relative gene expression levels were determined in this manner in each patient and the mean expression levels in the comparison populations were then calculated.

#### Fritz

[0220] Probe set MBXHUMMUS12164 detects Fritz nucleic acid sequences. Expression of Fritz transcripts was decreased in obese compared to lean patients in the gene profiling analysis.

B/C	Lean Pre Trog			Obese Pre Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	725	186	7	168	45	7	0.23	0.024	Fritz

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[0221] Expression of Fritz was also evaluated using real time PCR. The results further show that Fritz expression is significantly reduced in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (7) / Lean (7)	0.18	0.019

5 Legend "Fold change" indicates fold change in Fritz expression calculated as the ratio of mean obese expression/mean lean expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0222] Fritz contains the following protein domains (designated with reference to SEQ ID NO:2):

10 Fz domain encoded by amino acids 25-148,  
NTR/C345C module encoded by amino acids 186-295.

### PAK1B

15 [0223] Probe set MBXHUMMUS04063 detects p21 activated kinase 1B (PAK1B) nucleic acid sequences. Expression of PAK1B transcripts was increased in obese compared to lean patients in the gene profiling analysis.

B/C	Lean Pre-Trog			Obese Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	430	46	17	659	87	16	1.53	0.030	p21 activated kinase 1B (PAK1B)

20 Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[0224] PAK1B expression was also evaluated by real-time PCR. The results show that PAK1B is significantly over-expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (16) / Lean (17)	1.49	0.009

25 Legend "Fold change" indicates fold change in PAK1B expression calculated as the ratio of mean obese expression/mean lean expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

30 [0225] PAK1B and PAK1B splice variant contain the following protein domains (designated with reference to SEQ ID NO:6 and SEQ ID NO:8, respectively):

PBD domain encoded by amino acids 75 to 135,

Serine/threonine protein kinase domain encoded by amino acids 270 to 521.

[0226] PAK1B new splice variant contains the following protein domains (designated with reference to SEQ ID NO:10):

PBD domain encoded by amino acids 74 to 132.

5 Serine/threonine protein kinase domain encoded by amino acids 189 to 425.

**Serine Protease, umbilical vein endothelium (SPUVE)**

[0227] Probe set MBXHUMMUS14656 detects SPUVE nucleic acid sequences.

Expression of SPUVE transcripts was increased in obese compared to lean patients in the  
10 gene profiling analysis.

B/C	Lean Pre-Trog			Obese Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Exp	SEM	n			
B	502	35	7	905	149	7	1.80	0.036	SPUVE

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

15 [0228] SPUVE expression was also evaluated using real time PCR. The results further show that SPUVE is significantly over-expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (7) / Lean (7)	1.55	0.145

20 Legend "Fold change" indicates fold change in SPUVE expression calculated as the ratio of mean obese expression/mean lean expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0229] SPUVE contains the following protein domain (designated with reference to SEQ ID NO:16): trypsin domain encoded by amino acids 147-179.

25 **NK4**

[0230] Probe set MBXHUMMUS32474 detects NK4 nucleic acid sequences. Expression of transcripts encoding NK4 was increased in obese compared to lean patients in the gene profiling analysis.

B/C	Lean Pre-Trog			Obese Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	619	232	7	2020	419	7	3.26	0.017	NK4

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

5

[0231] NK4 expression was also evaluated using real time PCR. These data further show that NK4 is significantly over-expressed in muscle from obese individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Obese (16) / Lean (17)	3.22	0.004

10 Legend "Fold change" indicates fold change in NK4 expression calculated as the ratio of mean obese expression/mean lean expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

### Protein C Inhibitor (PCI)

15 [0232] Probe set Mbxhummus24898 detects expression of PCI nucleic acid sequences. Expression of transcripts encoding PCI was increased in diabetic patients compared to lean, non-diabetic patients in the gene profiling analysis described above.

B/C	Lean Pre-Trog			Diabetic Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	294	84	17	930	210	19	3.17	0.013	Protein C inhibitor

[0233] In addition, expression of transcripts encoding PCI was reduced in tro-treated patients compared to untreated patients.

B/C	Diabetic Pre-Trog			Diabetic Post-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	1422	340	9	722	247	10	0.51	0.005	Protein C inhibitor

20 Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetic post tro/diabetic pre tro.

25 [0234] PCI expression was also evaluated by real time PCR in lean vs. obese individuals and in pre-troglitazone-treated vs post-troglitazone-treated patients.

Comparison	Expression Fold change	t test
Diabetes (19)/Lean (17)	2.22	0.027
Diabetics + Tro (9) /Diabetics - Tro (9)	0.33	0.039

Legend "Tro" indicates treatment with Troglitazone. "Fold change" indicates fold change in TIEG expression calculated as either the ratio of mean diabetic pre-Tro expression/mean lean pre-Tro expression or mean diabetic post-Tro expression/mean diabetic pre-Tro expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0235] These data further demonstrate that PCI is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals. Furthermore, the data also show that treatment with Troglitazone reduces the expression of PCI in muscle from diabetic individuals.

[0236] Protein C inhibitor contains the following protein domains (designated with reference to SEQ ID NO:22): Signal Peptide at amino acids 1 to 16; and Serpin Domain at amino acids 50 to 406.

15

### MAST205

[0237] Probe set MBXHUMMUS12940 detects both MAST205 and MAST205b transcripts.

#### *Example 1*

[0238] This example shows that MAST205/MAST205b sequences are up-regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals in the gene profiling experiment described above.

B/C	Lean Pre-Trog			Diabetic Pre-Trog			Fold Change	Student's t test	Gene name
	Mean Exp	SEM	n	Mean Exp	SEM	n			
B	505	127	17	1086	131	19	2.15	0.003	MAST205

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetic post tro/diabetic pre tro.

#### *Example 2*

[0239] This example shows that MAST205/MAST205b are down-regulated in muscle of diabetics treated for 3 months with troglitazone in the gene profiling experiment described above.

B/C	Diabetic Pre-Trog			Diabetic Post-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	N	Mean Expr	SEM	N			
B	1234	411	9	919	325	8	0.74	0.01	MAST205

*Example 3*

[0240] Real-time PCR analysis further shows that MAST205 is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals.

5

Comparison	Expression Fold change	t test
Diabetes (19) / Lean (17)	1.45	0.001

Legend "Fold change" indicates fold change in MAST205 expression calculated as the ratio of mean obese expression/mean lean expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR

10 *Example 4*

[0241] This example shows that MAST205b is up regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals. It also demonstrates that MAST205b is down-regulated in muscle of diabetics after 3 months of troglitazone treatment compared to before treatment.

15 [0242] PCR primers and Taqman Probe were designed to detect specifically the expression of MAST205b. The sequences of the primers was as follows:

Forward primer: 110F – ACAGCAGTCCTGGCACTCCTT

Reverse primer: 174R – GCGGTTACTTGTCCGACAACTC

Probe133: TCCAGCCGCCACTGCCG

20

Lean Pre-Trog	Lean Post-Trog	Diabetic Pre-Trog	Diabetic Post-Tro			
Relative Exp (%)	Relative Exp (%)	Relative Exp (%)	Relative Exp (%)	Fold Change (D-/L-)	Fold Change (D+/D-)	Gene name
100	88	197	111	1.97	0.56	MAST205b

Legend: "Pre-Trog" and "Post-Trog" refer to samples taken before and after 3 months of troglitazone treatment respectively. "Relative Exp" refers to the expression of the gene relative to the Lean Pre-Trog sample, which is set to 100%. D-/L- refers to the ratio of relative expression in Diabetic Pre-Trog to relative expression in Lean Pre-Trog. D+/D- refers to the ratio of relative expression in Diabetic Post-Trog compared to relative expression in Diabetic Pre-Trog.

25

*Example 5*

[0243] This example shows that MAST205 is up regulated in muscle of diabetics when compared to muscle of lean non-diabetic individuals. It also demonstrates that MAST205 is

down-regulated in muscle of diabetics after 3 months of troglitazone treatment compared to before treatment.

[0244] PCR primers and Taqman Probe were designed to detect specifically the expression of MAST205. The sequences of the primers was as follows:

5 Forward primer: 717F – TTGGACAGTCTGCACCTTCTCTTA  
 Reverse primer: 801R – CGGTTACTTGTCCGACAAAAGC  
 Probe745: TGGCCTGAAGGACTTGAGCCTCCAGCCCCACTGCCG

Lean Pre-Trog	Lean Post-Trog	Diabetic Pre-Trog	Diabetic Post-Tro			
Relative Exp (%)	Relative Exp (%)	Relative Exp (%)	Relative Exp (%)	Fold Change (D-/L-)	Fold Change (D+/D-)	Gene name
100	98	242	66	2.42	0.27	MAST205

10 Legend: "Pre-Trog" and "Post-Trog" refer to samples taken before and after 3 months of troglitazone treatment respectively. "Relative Exp" refers to the expression of the gene relative to the Lean Pre-Trog sample, which is set to 100%. D-/L- refers to the ratio of relative expression in Diabetic Pre-Trog to relative expression in Lean Pre-Trog. D+/L+ refers to the ratio of relative expression in Diabetic Post-Trog compared to relative expression in Diabetic Pre-Trog.

15 *Example 6*

[0245] This example shows that MAST205 is up-regulated in skeletal muscle of DBA/2J mice fed a high fat diet. These mice became insulin resistant after 28 weeks on a 32% or 42% fat diet, compared to littermates fed a chow diet, as measured by IPIST.

	Chow Diet	32% Fat Diet	42% Fat Diet	Gene name
Mean Rel Exp (%)	118	153	170	Mouse MAST205
SEM	13	13	9	
N	5	5	5	
Fold Change	-	1.3	1.4	
Students T-test	-	0.09	0.008	

20 Legend: "Chow Diet" refers to standard mouse feed. "32% Fat Diet" and "42% Fat Diet" refer to mouse feed from in 32% or 42% of the calories in the diet are obtained from fat, respectively. "Mean Rel Exp (%)" refers to the average expression of the gene in muscles from 5 individual mice, relative to the expression in the muscle of a single mouse in the chow diet group.

25 colon Kruppel-like factor (CKLF)

[0246] Probe set MBXHUMMUS28900 detects CKLF nucleic acid sequences. Expression of transcripts encoding CKLF was higher in diabetic patients as compared to lean, non-diabetic patients in the gene profiling analysis described above.

B/C	Lean Pre-Trog			Diabetic Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	452	47	17	792	81	19	1.75	0.006	CKLF

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold change" indicates fold change of diabetic pre-trog/lean pre-trog.

5

[0247] Expression of transcripts encoding CKLF was also higher in obese patients as compared to lean patients in the gene profiling experiment.

B/C	Lean Pre-Trog			Obese Pre-Trog			Fold Change	Students t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	321	46	7	753	104	7	2.34	0.005	CKLF

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" indicates sample was

10 taken before 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold change" indicates fold change of obese pre-trog/lean pre-trog.

[0248] CKLF expression was also analyzed in diabetic individuals and lean individuals  
15 using real time PCR. The results showed that CKLF is significantly over-expressed in muscle from diabetic individuals when compared to muscle from lean individuals.

Comparison	Expression Fold change	t test
Diabetes (19) / Lean (17)	1.69	0.007

Legend "Fold change" indicates fold change of mean diabetic/mean lean.. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

20 [0249] Transcripts encoding CKLF contain the following protein domains (designated with reference to SEQ ID NO:34):

Zinc finger, C2H2 type domain encoded by amino acids 373 to 397

Zinc finger, C2H2 type domain encoded by amino acids 402 to 427

Zinc finger, C2H2 type domain encoded by amino acids 433 to 455.

## TABLE OF SEQUENCES

## SEQ ID NO:1 Human Fritz nucleic acid sequence

HUM161414 accession:U91903 coding sequence:70..1047

CGGAGACGGCGGAGCGGGCTTGGCGTCCACTGCGCGNTGCACCCCTGCCCATCCTGCCGGAT  
 5 CATGGTCTGCGGCAGCCGGGAGGGATGCTGCTGCGGGCTGCTGCCCTGGCTGCTCTGC  
 GCCTGCTCCGGGTGCCCGGGCTGGGCTGAGCCTGTGAGCCCGTCCGCATCCCCCTGTGCAAGTCC  
 CTGCCCTGGAACATGACTAAGATGCCAACCACTGCACCACAGCACTCAGGACAACGCCATCCTGGC  
 CATCGAGCAGTTCGAAGGTCTGCTGGGACCCACTGCAGCCCCGATCTGCTCTTCCCTGTGCCA  
 TGTACGCGCCCCTGCACCATTGACTTCCAGCACGCCATCAAGCCCTGTAAGTCTGTGCGAG  
 10 CGGGCCCGGCAGGGCTGTGAGCCATACTCATCAAGTACGCCACTCGTGGCCGGAGAACCTGCCCTG  
 CGAGGAGCTGCCAGTGTACGACAGGGCGTGTGCATCTCTCCGAGGCCATCGTTACTGCGGACGGAG  
 CTGATTTCTATGGATTCTAGTAACGAAACTGTAGAGGGGCAAGCAGTGAACGCTGTAATGTAAG  
 CCTATTAGAGCTACACAGAACCTATTCCGAAACAATTACAACATGTCAATTGGGCTAAAGTTAA  
 AGAGATAAAAGACTAAGTGCATGATGTGACTGCAGTAGTGGAGGTGAAGGAGATTCTAAAGTCCTCTC  
 15 TGGTAAACATTCCACGGGACACTGTCAACCTCTACCACTGGCTGCCTGCCCTCCACTTAAT  
 GTTAATGAGGAATATATCATCATGGCTATGAAGATGAGGAACGTTCCAGATTACTCTGGTGGAGG  
 CTCTATAGCTGAGAAGTGGAGGATCGACTCGTAAAAAGTTAAGCGCTGGGATATGAAGCTCGTC  
 ATCTGGACTCAGTAAAGTGATTCTAGCAATAGTGATTCCACTCAGAGTCAGAAGTCTGGCAGGAAC  
 TCGAACCCCCGGCAAGCACGCAACTAAATCCCAGAACACAGTGGACTTCCTATTAA  
 20 GACTTACTTGCATTGCTGGACTAGCAAAGGAAATTGCACTATTGCACATCATATTCTATTGTTACT  
 ATAAAAATCATGTGATAACTGATTATTACTTCTGTTCTCTGGTTCTGCTCTCTCA  
 ACCCCTTGTAATGGTTGGGGCAGACTCTTAAGTATATTGTGAGTTCTATTCACTAATCATGA  
 GAAAAACTGTTCTTGCAATAATAATAAAATTAACATGCTGTTACAGAGCCTCTTGCTGGAGTCT  
 CCAGATGTTAATTACTTCTGCACCCCAATTGGAATGCAATTGGATGAAAAGAGAGGTTCTGG  
 25 TATTACAGAAAGCTAGATATGCCCTAAACATACTCTGCCGATCTAATTACAGCCTATTGGTAT  
 GCCTTTGGCATTCTCCTCATGCTTAGAAAGTCCAAATGTTATAAGGTAATGGCAGTTGAA  
 GTCAAATGTCACATAGGCAAAGCAATCAAGCACCAAGGAAGTGTGTTATGAGGAAACACACCAAGATG  
 AATTATTTGAGACTGTCAGGAAGTAAATAAGGAGCTTAAGAAAGAACATTGCTGATTGA  
 GAAGCACAACTGAAACCAAGTAGCCGCTGGGTGTTAATGGTAGCATTCTCTGGCAATACATTG  
 30 ATTGTTCATGAATATATTAATCAGCATTAGAGAAATGAATTATAACTAGACATCTGCTGTTATCACC  
 ATAGTTTGTAAATTGCTCCTTAAATAACCCATTGGTGAAGTCCAAAAAAAAAAAAAAA  
 AAAAAA

## SEQ ID NO:2 Human Fritz polypeptide sequence

35 Protein sequence protein\_id:gi1917007

MVCSPGGMLLRAGLLALAALCLLRVPGARAAACEPVRIPLCKSLPWNMTKMPNHLHHSTQDNAILAIEQFEGL  
 LGTHCSPDLLFFLCAMYAPICTIDFQHEPIKPCSVCEARQGCEPILIKYRHSWPNLACEELPVYDRGVCISP

EAIVTADGADFPMDSSNGNCRGASSERCKCKPIRATQKTYFRNNNYVIRAKVKEIKTKCHDVTAVVEVKEILKS  
 SLVNIPRDTVNLYTSSGCLCPPLNVNEEYIIMGYEDEERSRLLLVEGSIAEKWKDRLGKKVKRWDMKLRHLGLSK  
 SDSSNSDSTQSQKSGRNSNPRQARN

5 **SEQ ID NO:3 Mouse Fritz nucleic acid sequence**

accession:NM\_011356 coding sequence: 365..1336  
 GAATTGGCACGAGCTGAATTGACTTTGTTTATTCTCTGGCTTCTCTGCCAGAGCGAGAGGAATAAATAGATGCTGCCAGCTGGAGCAGCTGGATCAA  
 10 GAGAACTGTGATTGTCAGGGTGGGGCAGCTCCCAGGTCGTTGGATCACCCCTCGGAACCGCAGGGGAG  
 ACTTCGGAACGAAAGTGTCTCCCGTCCGTGCGCCCTGCCATCCTGCTGGGACCATGGTCTGCG  
 CGGCCCCGGACGGATGCTGCTAGGATGGCCGGTTGCTAGTCCTGGCTGCTCTGCCAGGTGCCGG  
 AGCTCAGGCTGCAGCCTGTGAGCCTGTCGATCCGCTGTGCAAGTCCCTCCCTGGAACATGACCAAGATGCC  
 15 CAACCACTGCACCACAGCACCCAGGCTAACGCCATCTGCCATGGAACAGTCGAAGGGCTGCTGGCACCCA  
 CTGCAGCCCCGATCTCTCTTGCAATGTACGCACCCATTGCACCATCGACTTCCAGCACGAGCC  
 CATCAAGCCCTGCAAGTCTGTGAGCGCGCCGACAGGGCTGCGAGCCCATTCTCATCAAGTACGCCACTC  
 GTGGCCGGAAAGCTTGCCTGCGACGAGCTGCCGGTACGACCGCGGTGTCATCTCTCCGAGGCCATCGT  
 CACCGCGGACGGAGCGGATTTCTATGGATTCAAGTACTGGACACTGCAAGGGGCAAGCAGCGAACGTTGCAA  
 ATGTAAGCCTGTCAGAGCTACACAGAACCTATTCCGAAACAATTACAACATGTCACTCCGGCTAAAGTAA  
 20 AGAGGTAAAGATGAAATGTCATGATGTGACCGCCGTTGGAAGTGAAGGAATTCTAAAGGCATCACTGGTAAA  
 CATTCCAAGGGACACCGTCAATCTTATACCAACCTGGCTGCCCTGTCCTCCACTTACTGTCAATGAGGAATA  
 TGTCAATCATGGCTATGAAGACGAGGAACGTTCCAGGTTACTCTGGTAGAAGGCTCTAGCTGAGAAGTGGAA  
 GGATCGGCTTGGTAAGAAAGTCAAGCGCTGGATATGAAACTCCGACACCTGGACTGGTAAACTGATGCTAG  
 CGATTCCACTCAGAACATCAGAACGTTCTGGCAGGAACCTAACATCCCGGCCAGCACGAGCTAACATCTGAAATGAA  
 25 AAGGCCACACCCACGGACTCCCTCTAACAGACTGGCGCTGCTGGACTAACAAAGGAAACGCACAGTTGCTCGT  
 GACCGATTGTTACCGCAGACACCGCGTGGCTACCGAAGTTACTTCCGGTCCCTCTGCTTCTTAATGGC  
 CTGGGGTTAGATCTTAAATATGTTATATATTCTGTTCATCAATCACGTGGGACTGTTCTTGTCAACCCAGAA  
 TAGTAAATTAAATATGTTGATGCTAAGGTTCTGTACTGGACTCCCTGGTTAATTGGTGTCTGTACCTGA  
 TTGAGAATGCAATGTTCATGTAAGAGAGAACCTGGTCAATCTCAAGAACTAGATATTGCTGTAAGACAGCC  
 30 TCTGCTGCTGCTTATAGTCTTGTGTTGACCTGTTGGCCATTCCCTCATGCTGTGAAAGTTACATGTT  
 TATAAAGGTAGAACGGCATTGAAATCAGACACTGCACAAGCAGAGTAGGCCAACACCAGGAAGCATTATGAG  
 GAAACGCCACACAGCATGACTTATTTCAAGATTGGCAGGCAGAAAATAATAGTGGTGGAGCCAAGAAAAGA  
 ATATTTGCCCTGGTAAGGGCACACTGGAATCAGTAGCCTGAGCCATTACAGCAGTGTCTCTGGCACGTT  
 TTGATTTGTTCATAAATGTATTCAAGCAGCATTAGAGATGAACTTATAACTAGACATCTGTTCTACTATAG  
 35 CTCTGCTTCTCTAAATCAAACCCATTGTTGGATGCTCCCTCTCCATTCAAAATAATTGGCTTGCTGGTAT  
 TGCCAGGAAAAGAAAGTATTAAAGTATGCACTGCATGTGCAACAGGGTGTATTAAACAGAGGTATGTAACCTCA  
 TAAAAGACTATAATTACAGGACACGGAAATGTGCACATTGTTACTTTCTTCTTGTCTTGGCTTG  
 TGATTTGGTTGGTGTGTTATGTCATGTTGCTGTTAANCCATTGCACATTCAAGTTG  
 NACTAGATTAGAGTAGACTAGGCTCATTGCCCTAGACATTATGATTGAAATTGTTAATGCTCCATCAA  
 40 GATGTCTAATAAAAGGAATATGGTTGTCAACAGAGACGACAACACCCAAAAAAAAAAAAAA

**SEQ ID NO:4 Mouse Fritz polypeptide sequence**

Protein sequence accession:gi6755476

MVCCGPGRMLLGWAGLLVLAALCLLQVPGQAQAAACEPVRIPLCKSLPWNMTKMPNHLHHSTQANAILAMEQFEGL  
 5 LGTHCSPDLLFFLCAMYAPICTIDFQHEPIKPKSVCERARQCEPILIKYRHSWPESLACDELPVYDRGVCISP  
 EAIVTADGADFPMDSSSTGHCRGASSERCKCPVRATQKTYFRNNNYVIRAKVKEVVKMKCHDVTAVVEVKEILKA  
 SLVNI PRDTVNLYTTSGCLCPPLTVNEEYVIMGYEDEERSRLLLVEGSIAEKWKDRLGKKVKRWMKLRHLGLGK  
 TDASDSTQNQKSGRNSNPRPARS

**10 SEQ ID NO:5 Human PAK1B nucleic acid sequence**

HUM137707 accession:U51120 coding sequence:55..1692

GGAGAGCCGAGAGGAGCTGAGCGAGCGCGGAAGTAGCTGCTGGTGGTACAATGTCAAATAACGGCCTAGAC  
 ATTCAAGACAAACCCCCAGCCCCCTCCGATGAGAAATACCAGCACTATGATTGGAGCCGGCAGCAAAGATGCTGGA  
 15 ACCCTAAACCATGGTTCTAACCTCTGCCTCCAAACCCAGAGGGAGAAGAAAAAGAAGGACCGATTTACCGATCC  
 ATTTTACCTGGAGATAAAACAAATAAAAGAAAGAGAAAGAGCGGCCAGAGATTCTCTCCCTTCAGATTTGAA  
 CACACAATTCTATGTCGGTTTGATGCTGTACAGGGAGTTACCGGAATGCCAGAGCAGTGGGCCGCTTGCTT  
 CAGACATCAAATATCACTAAGTCGGAGCAGAAGAAAACCCGCAAGCTGTTCTGGATGTGGAGTTTACAAC  
 TCGAAGAAGACATCCAACAGCCAGAAATACATGAGCTTACAGATAAGTCAGCTGAGGATTACAATTCTCTAAT  
 GCCTTGAATGTGAAGGCTGTCTGAGACTCCTGCAGTGCACAGCTTCAGAAGATGAGGATGATGATGATGAT  
 20 GATGCTACCCCACCACCACTGATTGCTCCACGCCAGAGCACACAAATCTGTATACACACGGCTGTGATTGAA  
 CCACTTCCTGTCACTCCAACCTGGGACGTGGCTACATCTCCATTTCACCTACTGAAAATAACACCAACTCCACCA  
 GATGCTTGACCCCTTAATACTGAGAAGCAGAAGAAGCCTAAAATGTCTGATGAGGAGATCTGGAGAAATTAA  
 CGAACGATAGTGAGTGTGGCGATCTAAGAAGAAATATACACGGTTGAGAAGATTGGACAAGGTGCTTCAGGC  
 ACCGTGTACACAGCAATGGATGTGGCCACAGGACAGGGAGGTGGCCATTAAGCAGATGAATCTTCAGCAGCAGCCC  
 25 AAGAAAGAGCTGATTATAATGAGATCCTGGTCATGAGGGAAAACAAGAACCCAAACATTGTGAATTACTTGGAC  
 AGTTACCTCGTGGGAGATGAGCTGTGGTTATGGAATACTTGGCTGGAGGCTCCTTGACAGATGTGGTACA  
 GAAACTTGCATGGATGAAGGCCAATTGCAGCTGTGCCGTGAGTGTCTGCAGGCTCTGGAGTCTTGCAATTG  
 AACCAGGTCAATTACAGAGACATCAAGAGTACAATATTCTGTTGGAAATGGATGGCTCTGTCAAGCTAACTGAC  
 TTGGATTCTGTGCACAGATAACCCAGAGCAGAGCAAACGGAGCACCATTGGTAGGAACCCATACTGGATGGCA  
 30 CCAGAGGTTGTGACACGAAAGGCCATGGGCCAAGGTTGACATCTGGCCCTGGCATCATGGCCATCGAAATG  
 ATTGAAGGGGAGCCTCCATACCTCAATGAAAACCTCTGAGAGCCTTGTACCTCATTGCCACCAATGGACCCCA  
 GAACCTCAGAACCCAGAGAAGCTGTCAGCTATCTTCCGGACTTCTGAACCGCTGTGAGGATGGATGTGGAG  
 AAGAGAGGTTCAAGAGCTGCTACAGCATCAATTCTGAAGATTGCCAAGCCCCTCTCCAGCCTCACTCCA  
 CTGATTGCTGAGCTAAGGAGGCAACAAAGAACATCACTAAACACACTCACCCAGCCTCATTGTGCCAAGC  
 35 CTTCTGTGAGATAAA

**SEQ ID NO:6 Human PAK1B polypeptide sequence**

protein\_id:gi1256422

MSNNGLDIQDKPPAPPMRNTSTMIGAGSKDAGTLNHGSKPLPPNPEEKKKDRFYRSILPGDKTNKKKEKERPEI  
 5 SLPSDFEHTIHVGFDATGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKTSNSQKYSFTDKSA  
 EDYNSSNALNVKAVSETPAVPPVSEDEDDDDDDATPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT  
 ENNTTPPDALTLNTEQKKPKMSDEEILEKLRSIVSVGDPKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQ  
 MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGSILDVVTETCMDEGQIAAVCRECLQ  
 ALESLSNQVIHRDIKSDNILLGMDGSVKLTDGFCAQITPEQSKRSTMVGTPYWMAPEVVTRKAYGPKVDIWSL  
 GIMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQHQFLKIAK  
 10 PLSSLTPLIAAAKEATKNNH

**SEQ ID NO:7 Human PAK1B splice variant nucleic acid sequence**

accession:AF0718844      coding sequence:12..1673

TGGTGGTGACAATGTCAAATAACGGCCTAGACATTCAAGACAAACCCCCAGCCCCTCCGATGAGAAATACCAAGCA  
 15 CTATGATTGGAGTCGGCAGCAAAGATGCTGGAACCCCTAAACCATGGTTCTAAACCTCTGCCTCCAAACCCAGAGG  
 AGAAAGAAAAAGAAGGACCGATTTCACCGATCCATTACCTGGAGATAAAACAAATAAAAGAAAGAGAAAGAGC  
 GCCCAGAGATTCTCTCCCTTCAGATTGAAACACACAATTCAATGTCGGTTTGATGCTGTACAGGGGAGTTA  
 CGGGAATGCCAGAGCAGTGGGCCGCTTGCTTCAGACATCAAATATCACTAAGTCGGAGCAGAAGAAAACCCGC  
 AGGCTGTTCTGGATGTGGAGTTTACAACCTCGAAGAAGACATCCAACAGCCAGAAATACATGAGCTTACAG  
 20 ATAAGTCAGCTGAGGATTACAATTCTCTAATGCCCTGAATGTGAAGGCTGTGAGACTCCTGCAGTGCCAC  
 CAGTTTCAGAAGATGAGGATGATGATGATGATGCTACCCACCAGTATTGCTCCACGCCAGAGCACA  
 CAAAATCTGTATAACACACGGCTGTGATTGAACCACTTCCTGTCACTCCAACCTGGACGTGGCTACATCTCCA  
 TTTCACCTACTGAAAATAACACCACTCCACCAAGATGCTTGACCCGGAATACTGAGAACAGAGAAGAAGCCTA  
 AAATGTCTGATGAGGAGATCTTGAGAAATTACGAAGCATAGTGAGTGTGGCGATCTAAGAAGAAATATACAC  
 25 GGGTTGAGAAGATTGGACAAGGTGCTTCAGGCACCGTGTACACAGCAATGGATGTGGCCACAGGACAGGAGGTGG  
 CCATTAAGCAGATGAATCTTCAGCAGCAGCCAAAGAAAGAGCTGATTATTAATGAGATCCTGGTCATGAGGGAAA  
 ACAAGAACCCAAACATTGTGAATTACTTGGACAGTTACCTCGTGGAGATGAGCTGTGGGTGTTATGGAATACT  
 TGGCTGGAGGCTCCTTGACAGATGTGGTACAGAAACTTGCAATGGATGAAGGCCAAATTGCAGCTGTGCCGTG  
 AGTGTCTGCAGGCTCTGGAGTTCTGCATTCAACCAAGGTCAATTCAACAGAGACATCAAGAGTGACAATATTCTGT  
 30 TGGGAATGGATGGCTCTGTCAGCTAACTGACTTTGGATTCTGTCACAGATAACCCAGAGCAGAGCAAACCGGA  
 GCACCATGGTAGGAACCCATACTGGATGGCACCAGAGGTTGTACAGCAAAGGCCTATGGGCCAAGGTTGACA  
 TCTGGTCCCTGGCATCATGCCATCGAAATGATTGAAGGGAGCCTCCATACCTCAATGAAAACCCCTTGAGAG  
 CCTTGTACCTCATTGCCACCAATGGACCCAGAACTTCAGAACCCAGAGAACGCTGTCAAGCTATCTTCCGGACT  
 TTCTGAACCGCTGTCTGAGATGGATGTGGAGAAGAGAGGTTCAAGCTGCTACAGGTGAGAAAATCTGA  
 35 GGGTTCAAGTGTGTTAGTAACCTTCCATGATAGCTGCATCAATTCTGAAAGATTGCCAAGCCCTCTCCAGCCTC  
 ACTCCACTGATTGCTGCAGCTAAGGAGGCAACAAAGAACAAACTCACTAAACCAACTCACCCAGCCTCATTGTG  
 CCAAGCTCTGTGAGATAAAATGCACATTCAAGAAATTCCAACCTCTGATGCCCTTCTCCTTGCCCTGCTTCTCC  
 CATTTCCTGATCTAGCACTCCTCAAGACTTGTGATCCTGGAAACCGTGTCCAGCATTGAAGAGAACTGCAACT  
 GAATG

**SEQ ID NO:8 Human PAK1B splice variant polypeptide sequence**

protein\_id:gi3265160

MSNNGLDIQDKPPAPPMRNTSTMIGVGSKDAGTLNHGSKPLPPNPEEKKKDRFYRSILPGDKTNKKKEKERPEI  
 5 SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKTNSQKYMSTDKSA  
 EDYNSSNALNVKAVSETPAVPPVSEDEDDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT  
 ENNTTPPDALTRNTEKQKKPKMSDEEILEKRSIVSGDPKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQ  
 MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMYEYLAGGS LTDVVTETCMDEGQIAAVCRECLQ  
 ALEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQS KRSTMVGT PYWMAPEVVTRKAYGPKVDIWSL  
 10 GIMAIEMIEGEPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQVRKLRFQV  
 FSNFSMIAASIPEDCQAPLQPHSTDCCS

**SEQ ID NO:9 Human PAK1B new splice variant nucleic acid sequence**

TCAAATAACGGCCTAGACATTCAAGACAAACCCCCAGCCCCCTCCGATGAGAAATACCAGCACTATGAT  
 TGGAGCCGGCAGCAAAGATGCTGGAACCTAAACCATGGTCTAACCTCTGCCTCCAAACCCAGAGG  
 15 AGAAGAAAAAGAAGGACCGATTTACCGATCCATTTACCTGGAGATAAAACAAATAAAAGAAAGAG  
 AAAGAGCGGCCAGAGATTCTCCCTTCAGATTGAAACACACAATTCATGTCGGTTTGATGCTGT  
 CACAGGGGAGTTACGGGAATGCCAGAGCAGTGGCCGCTTGCTCAGACATCAAATATCACTAAGT  
 CGGAGCAGAAGAAAAACCGCAGGCTGTTCTGGATGTGTTGAGATTACAACCTCGAAGAACATCC  
 AACAGCCAGAAATACATGAGCTTACAGATAAGTCAGCTGAGGATTACAATTCTTAATGCCTTGAA  
 20 TGTGAAGGCTGTCTGAGACTCCTGCAGTGCCACCAGTTTCAAAGATGAGGATGATGATGATGATG  
 ATGCTACCCACCACCACTGATTGCTCCACGCCAGAGCACACAAATCTGGCCATTAAAGCAGATG  
 AATCTTCAGCAGCAGCCAAGAAAGAGCTGATTATTAATGAGATCCTGGTCATGAGGGAAACAAGAA  
 CCCAACATTGTAATTACTTGGACAGTTACCTCGTGGAGATGAGCTGTGGTTATGGAATACT  
 TGGCTGGAGGCTCCTTGACAGATGTGGTACAGAAACTGCATGGATGAAGGCCAAATTGCAGCTGT  
 25 TGCGTGAGTGTCTGCAGGCTCTGGAGTTCTGCATTGAAACCAGGTCAATTACAGAGACATCAAGAG  
 TGACAATATTCTGGATGGATGGCTCTGTCAGCTAACTGACTTGGATTCTGTCACAGATAA  
 CCCCAGAGCAGAGCAAACGGGACCCATGGTAGGAACCCATACTGGATGGCACCAGAGGTTGTGACA  
 CGAAAGGCCATGGGCCAAGGTTGACATCTGGCCCTGGCATCATGCCATCGAAATGATTGAAGG  
 GGAGCCTCCATACCTCAATGAAAACCTCTGAGAGCCTGTACCTCATTGCCACCAATGGACCCAG  
 30 AACCTCAGAACCCAGAGAAGCTGTCAGCTATCTCCGGACTTCTGAACCGCTGTCTGGATGGAT  
 GTGGAGAAGAGAGGTTCAGCTAAAGAGCTGCTACAGCATCAATTCTGAAGATTGCCAAGGCCCTCTC  
 CAGCCTCACTCCACTGATTGCTGCAGCTAAGGAGGCAACAAAGAACATCACTAA

**SEQ ID NO:10 Human PAK1B new splice variant polypeptide sequence**

35 MSNNGLDIQDKPPAPPMRNTSTMIGAGSKDAGTLNHGSKPLPPNPEEKKKDRFYRSILPGDKTNKKK  
 EKERPEISLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKT  
 SNSQKYMSTDKSAEDYNSSNALNVKAVSETPAVPPVSEDEDDDDATPPPVIAPRPEHTKSVAIKQ

MNLQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMEYLAGGS LTDVVTETCMDEGQIAA  
VCRECLQALEFLHSNQVIHARDIKSDNILLGMDGSVKLTDFGFCAQITPEQSKRSTMVGTPYWMAPEVV  
TRKAYGPKVDIWSLGIMAIEMIEGEPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLGM  
DVEKRGSAKELLOHOFLKIAKPLSSLTPLIAAAKEATKNNH

5

**SEQ ID NO:11 Mouse PAK1B nucleic acid sequence**

accession:NM\_011035 coding sequence:190..1827

coding sequence: 190..1827

10 TGAGTGTGCTGCTCCTGGGACGTGGGTGGCTGCGCTGCGCTGTTAGCCGAATCCAGCCTGTCAGGCTGGTTCA  
GCCCGGGCTTGTCCAAGTGTGAGCAGGGAGAGGTGCGGCTCTCGAGACTCACAGATAACAAAGATCACGCC  
CGCACCCACCGCCAGTAGCTGCTGCTGGTGGTACAATGTCAAATAACGGCGTAGACATCCAGGACAAACCC  
CCAGCCCCCTCCGATGAGAAACACCAGCACTATGATTGGAGCCGGCAGCAAAGACACTGGAACCTAAACCACGGC  
TCCAAACCTCTGCCTCCAAACCCAGAGGGAGAAGAAAAGAAGGACCGGTTTATCGATCCATCTTACCTGGAGAT  
AAAACAAATAAAAAGAGGGAGAAGGAGCGACCAGAGATTTCTTCAGATTGAGCATACAATTGATGTT  
GGTTTGATGCTGTACAGGGAGTTACGGGAATGCCAGAACAGTGGGCTCGCTGCTCAAACATCAAATATC  
15 ACAAAAGTCAGAGCAGAAGAAAAACCCACAGGCTGTTCTGGATGTGTTGAATTTATAACTCTAAGAAGACCTCC  
AATAGTAAGAAGTACATGAGTTTACAGATAAGTCAGCTGAAGATTATAATTCTCTAACACTTGAATGTGAAG  
ACTGTGTCTGAGACCCCAGCAGTACCAACCAGTGTAGAAGATGATGATGACGATGCTACCCACCT  
CCAGTGATTGCTCCAGGCCAGAACACACAAAATCTGTATATACACGATCTGTGATTGAACCACTCCCTGTTACT  
CCAACTCGGGATGTGGCTACATCTCCTATTCTCCTACTGAGAATAACACCACTCCGCCAGATGCTTGACCCGG  
20 AACACGGAAAACAGAAGAAGAAGCCTAAAATGTCGTGAGGGAGATCTAGAGAAATTACGGAGCATAGTGAGT  
GTGGGTGACCCCCAAGAAGAAGTACACACCGTTGAGAAGATTGGACAAGGTGCTTCAGGCACAGTGATACTGCA  
ATGGATGTAGCCACAGGCAGGGAGGTGGCATTAAACAGATGAATCTCAGCAGGCCAGAACAGCTGATT  
ATTAATGAGATCCTGGTCATGAGGGAAAACAAAACCCAAATATTGTCAACTACCTGGACAGTTACCTGTGGGA  
GATGAGCTGTGGTTGTTATGGAATACTTGGCTGGAGGCTCTGACAGATGTGGTACAGAAACCTGTATGGAT  
25 GAAGGCCAGATAGCAGCTGTGCGAGAGTGTCTACAAGCTTGGAGTTCTACATTCAAACCAAGTCATTCA  
AGGGACATCAAGAGTGACAATTCTGCTGGGAATGGATGGCTCTGTCAAGTTAAGTCACTTGGATTCTGTGCA  
CAGATAACTCCAGAGCAGAGCAAAAGGAGCACCAGGTGGGAACCTCCATATTGGATGGCACCTGAAGTGTGACA  
CGCAAGGCTTATGGACCCAAGGTTGACATCTGGTCCCTGGCATTATGCAATTGAAATGATTGAGGGGGAGCCC  
CCATACCTCAATGAAAACCTTGTAGAGGCCCTGTACCTCATTGCTACCAATGGGACGCCAGAGCTTCAGAACCC  
30 GAGAAGTTGTCAGCTATCTTCCGGGACTTCTGCAATGCTGTTGAGATGGATGTGGAGAACAGAGAGGCTCAGCT  
AAAGAGCTGCTGCAGCATCAGTTCTGAAGATTGCAAGGCCCTCTAGCCTGACTCCACTGATGCTGCA  
AAAGAGGCAACCAAGAACAACTCACTGAAACCAGCTCATCCAGCCTCATGTGCAAGCCTTATGAAATAAAC  
ACTTGTTCGGGAACTCCGACACCTCATGTCCTCTCCTTCCATTGCTTCTCCATTGATCTAGTGCTC  
CCAAGACTTGTATCCTGGAAACTGTC

35

**SEQ ID NO:12 Mouse PAK1B polypeptide sequence**

accession:gi6754978

MSNNGVDIQDKPPAPPMRNTSTMIGASKDTGTLNHGSKPLPPNPEEKKKDRFYRSILPGDKNKKREKERPEI  
SLPSDFEHTIHVGFDAVTGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSKKTSNSKKYMSFTDKSA

EDYNSSNTLNVKTVSETPAVPPVSEDDDEDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPT  
 ENNTTPPDALTRNTEKQKKPKMSDEEILEKRSIVSGDPKKYTPFEKIGQGASGTVYTAMDVATGQEVAIKQ  
 MNLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMYEYLAGGS LTDVVTETCMDEGQIAAVCRECLQ  
 5 ALEFLHSNQVIHRDIKSDNILLGMDGSVKLTDFGFCAQITPEQS KRSTMVGT PYWMAPEVVTRKAYGPKVDIWSL  
 GIMAIEMIEGEPPYLNEPLRALYLIATNGTPELQNPEKLSAIFRDFLQCCLMEDVEKRGSAKELLHQFLKIAK  
 PLSSLTPLMHAKEATKNNH

**SEQ ID NO:13 Rat PAK1B nucleic acid sequence**

accession:NM\_017198 coding sequence:389-2023

10 CCCGGGCGGAAGGTACCGCCGACGGTGGCGCTCCTCACCGCTTCCCCACACCTCCCAGTCCCCTGCACCTCT  
 AGGCTGCGCGAGTCCTCTCGGCTATTACCGGCTCTTGGAGCCGCTGTGCCCCCTCAGAGCCCGCGCTCTCCA  
 CAGTCCTCAGAGAAGGACCCCTCAGAGAGCGCACCCGCCCTCGCAGTCGGCGCGCCTCCGCTCCCC  
 GCGGCCCGCAGCGTCCGCCCTCAGAGCCGGCTCAGAGCCGGAGAGAGAACGCGGGCTGCCGCCGCGCAC  
 ACTCCCGCGCTCTCTGGGTTGGCTGCCGCCAGGAGGAGGAGCGAGCGAGCGAGAAGTAGCTGCTGCT  
 15 GCTGGTGGTACAATGTCAAATAACGGCTTAGACGTCCAGGACAAACCCAGCCCCCTCCGATGAGAAACACCAG  
 CACTATGATTGGAGCTGGCAGCAAAGACCCCTGGAACCTAAACCACGGTTCAAACCTCTGCCCTCAAACCCAGA  
 GGAGAAGAAAAAGAAGGACCGTTCTATCGATCCATCTTAGCTGGAGATAAAACAAATAAAAGAAGGAGAAGGA  
 GCGGCCAGAGATTCTTCCTTCAGATTGAGCATAACATTCTAGTGGTTTGATGCTGTACAGGGAGTT  
 TACGGGGATGCCAGAACAGTGGGCCGCTTGCTCAAACATCAAATATCACCAGTCAGAGCAGAACAAAAACCC  
 20 ACAGGCTGTTCTGGATGTGGAAATTATAACTCCAAGAACCTCCAATAGTCAGAAGTACATGAGTTTAC  
 AGATAAGTCAGCTGAAGATTATAATTCTCTAATACTTGAATGTGAAGACTGTGTCTGAGACCCAGCAGTGCC  
 ACCAGTGTCAAGAGATGAAGATGATGATGACGATGCTACCCACCTCCAGTGATTGCTCCAGGCCAGAACACAC  
 AAAATCTGTATATACACGGTCTGTGATTGAACCACCTCCTGTTACTCCAACTCGGATGTGGCTACATCTCTAT  
 TTCTCTACTGAGAATAACACCACTCCGCCAGATGCTTGACCCCGAATACTGAAAAGCAGAAGAAGCCTAA  
 25 AATGTCTGATGAGGAGATCTTAGAGAAATTACGGAGCATAGTGAGTGTGGCGATCTAAGAAGAAGTACACACG  
 CTTCGAGAAGATTGGACAAGGTGCTTCAGGCACAGTGTACACTGCAATGGATGTAGCCACAGGGCAGGAGGTGGC  
 CATTAAACAGATGAACCTTCAGCAGCAGCCGAAGAAAGAACTCATTATTAAATGAGATCCTGGTATGAGGGAAAA  
 CAAAAACCCAAACATTGTGAACTATCTGGACAGTTACCTGTGGAGATGAGCTATGGTTGTATGGAATACTT  
 GGCGGGAGGCTCTAACAGATGTGGTACAGAACCTGTATGGATGAAGGCCAGATAGCAGCTGTGCGGAGA  
 30 GTGTCTACAAGCTTGGAGTTCCTGCATTCAAACCAAGTCATTACAGAGACATCAAGAGTGACAATATTCTGCT  
 GGGAAATGGATGGCTCTGTCAAATTAACTGACTTGGATTCTGTGCACAGATAACTCCAGAGCAGAGCAAAAGGAG  
 CACCATGGTGGAACTCCATATTGGATGGCACCTGAAGTTGTGACACGCAAGGCCTATGGACCCAAGGTTGACAT  
 CTGGTCCCTGGTATTATGGCAATTGAAATGATTGAGGGGGAGCCCCCATACCTCAATGAAAACCTTTGAGAGC  
 CTTGTACCTCATTGCTACCAATGGACGCCAGAGCTTCAGAACCCAGAGAAGTTGTCAAGCTATTTCGGGACTT  
 35 TTTGAACCGCTGTCTTGAGATGGATGTGGAGAAGAGAGGTTCAAGCTAAAGAGCTACTGCAGCAGCATCAATTCTGAA  
 GATTGCCAAGCCTCTCCAGCCTGACTCCACTGATTGCTGCAGCAAAAGAGGCAACCAAGAACATCACTGAAA  
 CCACGCTCACCCAGCCTCATGTGCCAAGCCTCTATGAAATAAACACTCGTTGGAAACTCCGACCCCTCATG  
 TCCTCTTCTCCTTCCCTGCTCTCCATTCCGATCTAGTGCTCTCAAGACACTTGTGATCCTGGAAACCGTCTA  
 GCACTGAAGAGAACCGCAACCGGATGACTAATCCAGCAGAGGCCATTCTAAATAGGAATTCCCTTAGCTTG  
 40 GGCATGGAGGGACTGATGAGCAAGGGTTACCTGAATAAACCTGTTACGAAACAGAAATCTCAACCACCC

ATTCTTACCCCTACAATCAGTCTTAACTCTATAAACTTATGGTTGATAGCATTATCAATTGCTATCAGTT  
 GAAATTGCTTTGTTTTATTCTGTGACCAAATTGCCAACACTTCATTGTATTGAAAACCAGAACAGCTT  
 TGAAATGCCACGGGCCTGATAATCTGCCAGGGACATGAAGAGGTCTTGTTCCTGAACCCAC

5 **SEQ ID NO:14 Rat PAK1B polypeptide sequence**

accession:gi8393901

MSNNGLDVQDKPPAPPMRNTSTMIGASKDPGTLNHGSKPLPPNPEEKKKDRFYRSILAGDKTNKKKEKERPEI  
 SLPDFEHTIHVGFDATGEFTGMPEQWARLLQTSNITKSEQKKNPQAVLDVLEFYNSSKTSNSQKYMSTDKSA  
 EDYNSNTLNVKTVSETPAVPPVSEDEDDDDATPPPVIAPRPEHTKSVYTRSVIEPLPVTPTRDVATSPISPTE  
 10 NNTTPPDALTRNTEKQKKPKMSDEEILEKLRSIVSVGDPKKYTRFEKIGQGASGTVYTAMDVATGQEVAIKQM  
 NLQQQPKKELIINEILVMRENKNPNIVNYLDSYLVGDELWVVMELYLAGGSLTDVVTECMDEGQIAAVCRECLQA  
 LEFLHSNQVIHRDIKSDNILLGMDGSVKLTDGFCAQITPEQSFRSTMVGTWMAPEVVTRKAYGPKVDIWSLG  
 IMAIEMIEGEPPYLNENPLRALYLIATNGTPELQNPEKLSAIFRDFLNRCLEMDVEKRGSAKELLQHQFLKIAKP  
 LSSLTPLIAAAKEATKNNH

15

**SEQ ID NO:15 Human SPUVE nucleic acid sequence**

HUM168767 accession:BC001278 coding sequence:121..1272

GACCCACGCGTCCGGGGCAGGCATGGAGCCGCGCCTCTCTCCCGCGCCCACACCTGTCTGAGCGGCGCAGC  
 GAGCCGCGGCCCGGGCGGCTGCTCGCGCGGAACAGTGCTCGGCATGGCAGGGATTCCAGGGCTCCTCTTCTT  
 20 CTCTTCTTCTGCTCTGTGCTTGGCAAGTGAGCCCTTACAGTGCCCCCTGAAACCCACTTGGCCTGCATAC  
 CGCCTCCCTGCGCTTGCCCCAGTCTACCCCTCAATTAGCCAAGCCAGACTTGGAGCCGAAGCCAAATTAGAA  
 GTATCTCTTCATGTGGACCCCAGTGTCTAAGGAACTCCACTGCCACTTACGAAGAGGCCAAGCAATATCTG  
 TCTTATGAAACGCTCTATGCCATGGCAGCCGACAGAGACGCAGGTGGCATCTACATCCTCAGCAGTAGTGG  
 GATGGGGCCAACACCGAGACTCAGGGTCTCAGGAAAGTCTCGAAGGAAGCGGCAGATTATGGCTATGACAGC  
 25 AGGTTCAGCATTGGAGGACTTCCTGCTCAACTACCCCTTCAACATCAGTGAAGTTATCCACGGGCTGC  
 ACCGGCACCCCTGGTGGCAGAGAACATGTCTCACAGCTGCCACTGCATACACGATGGAAAAACCTATGTGAA  
 GGAACCCAGAACGCTCGAGTGGCTTCTAAAGCCAAGTTAAAGATGGTGGCGAGGGGCCAACGACTCCACT  
 TCAGCCATGCCGAGCAGATGAAATTTCAGTGGATCCGGTGAAACGCACCCATGTGCCCAAGGGTTGGATCAAG  
 GGCATGCCATGACATCGGCATGGATTATGATTATGCCCTCTGGAACTCAAAAGCCCCACAAGAGAAAATT  
 30 ATGAAGATTGGGTGAGCCCTCTGCTAAGCAGCTGCCAGGGGCAGAAATTCACTTCTCTGGTTATGACAATGAC  
 CGACCAGGCAATTGGTGTATCGCTCTGTGACGTCAAAGACGAGACCTATGACTTGCCTACCAAGCAATGCGAT  
 GCCCAGCCAGGGGCCAGCGGGCTGGGTCTATGTGAGGATGTGGAGAGACAGCAGCAGAACGTGGAGCGAAAA  
 ATTATTGGCATTTCAGGGCACCAGTGGTGGACATGAATGGTCCCCACAGGATTCAACGTGGCTGTCAGA  
 ATCACTCCTCTCAAATATGCCAGATTGCTATTGGATTAAAGGAAACTACCTGGATTGTAGGGAGGGGTGACAC  
 35 AGTGTCCCTCCTGGCAGCAATTAAAGGTCTTCATGTTCTTATTAGGAGAGGCCAAATTGTTTGTCAATTG  
 GCGTGCACACGTGTGTGTGTGTGTGTAAAGGTGCTTATAATCTTACCTATTCTTACAATTGC  
 AAGATGACTGGCTTTACTATTGAAAAGTGGTTGTATCATATCATATCAATTAAAGCAGTTGAAGGCATA  
 CTTTGCAATAGAAATAAAAAAAACTGATTGGGCAATGAGGAATATTGACAATTAAAGTTAATCTTCACGTT  
 TTTGCAAACTTGATTTCATCTGAACCTGTTCAAAGATTATTAATATTGGCATAACAAGAGAAA  
 40 AAAAAAA

**SEQ ID NO:16 Human SPUVE polypeptide sequence**

Protein sequence protein\_id:gi12654869

MAGIPGLLFLLFFLLCAVGQVSPYSAPWKPTWPAYRLPVVLQSTLNLAKPDFGAEAKLEVSSSGPQCHKGTP  
PTYEEAKQYLSEYETLYANGSRTETQVGIVYILSSSGDGAQHRDSSGKSRKRQIYGYDSRFSIFGKDFLLNYPF  
STSVKLSTGCTGTLVAEKHLTAACIHDKTYVKGTQKLRVGFLKPKFKDGGRGANDSTSAMPEQMKFQWIRVK  
RTHVPKGWIKGNANDIGMDYDYALLELKKPHKRKFMKIGVSPPAKQLPGGRIHFSGYDNDRPGNLVYRFCDVKDE  
TYDLLYQQCDAQPGASGSGVYVRMWKRQQQKWERKIIGIFSGHQVDMNGSPQDFNVAVRITPLKYAQICYWIKG  
NYLDCREG

10

**SEQ ID NO:17** Mouse SPUVE nucleic acid sequence

accession:BC018517 coding sequence:170..1318

15 GTGCTGGTGTCACTAGACTCTTCTGGCAGGGTACCTTCATGACTCTGCCTTATGAGAAGATTCTCAGAAC  
TGACACCCCTCCGGAGCCCAGGGACAGAGCTGCAGGAGAGGCCTGCCTGCCTGCCTGCATGCCGCC  
TTCCCTATGTAGCACTCAGCATGGCTGGAATCCGGGCTCTTCATCCTTCTTGTCCCTGCCTGTGTGTTCATGCA  
GGTAGTCCCTACACCGTTCCGTGGAAACCCACATGGCCGGCTTATGCCCTCCCTGTAGTCTGCCTCAGTCTAC  
CCTCAACTTAGCTAAGGCAGACTTCGACGCCAAGCGAAATTGGAGGTGCCTCCTCATGTGGACCTCAGTGTCA  
CAAGGGAACACCACTGCCACCTACGAAGAGGCCAAGCAGTACCTTCCATGAAACCCTTATGCCAATGGCAG  
CCGCACAGAGACTCGGGTGGGATCTACATCCTCAGCAATGGTGAAGGCAGGGCACGGCAGAGACTCGGAGGC  
20 CACAGGGAGATCTCGCAGGAAGAGGCAGATTATGGCTACGATGGCAGGTTAGCATTGGAGGACTTCCT  
GCTCAATTATCCTTCTCAACATCGGTGAAGTTGTCTACTGGCTGCAGTGGCACCCCTGGTGGCAGAGAACGACGT  
CCTCACTGCTGCCACTGCATACACGATGGAAAACCTATGTGAAAGGGACACAGAAAACCTCCGAGTGGCTTCC  
GAAGGCCAAGTATAAAGATGGTGCAGGGGACAAACAGCTCGAGCTCAGCCATGCCAGACAAGATGAAGTTCA  
GTGGATCCCGTGAAACGCACCCATGTGCCAAGGGGTGGATCAAGGGCAATGCCAATGACATCGGATGGATT  
25 TGACTACGCCCTGCTGGAACTCAAGAAACCCACAAAAGACAGTTCATGAAGATTGGTGTGAGTCCTCCAGCGAA  
GCAGCTCCAGGGGCAGGATCCACTTCTCTGGTTATGACAATGACCGGCCGGCAATTGGTGTACCGCTTCTG  
TGATGTCAAAGATGAGACCTACGACCTTCTTACCAAGCAGTGTGACGCCAGCCCCGGCCAGTGGTTAGGGGT  
CTATGTGAGGATGTGGAAGAGACCACAGCAGAAATGGAAAGAAAAATTATCGGCATTTTCAGGGCACCAAGTG  
GGTGGACATGAATGGCTCTCACAGGATTCAACGTGGCAGTTAGAATCACGCCCTTAAATATGCCAGATTG  
30 CTATTGGATTAAAGGAAACTACCTAGATTGCAGGGAGGGGTGACATGCGTCTTCTGCCAGCACCAATGGTCTTT  
TTGCACTCATTGTAGGAGAGGCTAGTTTTATCATTGACTCTTGTGGTGTGAGTCACATAGTATTTTACCTA  
GTATTCTCAAATGCCAAAATTATTGGCTATATTATTAAACTGTTGTGCGTTATAGCATTAAAGCAGTC  
TGAAAGCATACTTGCATAGAGACTTAAAGTATTGGTAATAGGGCTATTGACAAGGAAGTTAAACTTTC  
AGTTTTGGAGAATTCTAATTGGTCTGATCCAAACTGCTTCAGAGGTTATATCAAATACGTGACACACAGG  
35 GAATATGAATTCTTATGTTGTATATGTATATGTTCTTCTGAGAGTCATATATTGATATTGGTAATGTGTG  
GTTATTATGCTCCAGATAATGATAGCAGGAAAGTCTCAATAGGCAATTATAATGTTGGATTCAAACATTACG  
TAGTAGCCTTGAAGAGAACAAATAATTATTGGCTATATTGATACCCATATAAGACTGTATCTTACAGTCACAG  
AATTCCCAAGCCTGCTTTAGTTGAAAATAACCTTCCCTGTAAAAAAAAAAAAAAAAAAAAAA

**SEQ ID NO:18 Mouse SPUVE polypeptide sequence**

Protein sequence accession:gi20178034

MAGIPGLFILLVLLCVFMQVSPYTVPKPTWPAYRLPVVLQPSTLNLA  
 KADFDAAKAKLEVSSSCGPQCHKGTPLP  
 TYEEAKQYLSYETLYANGSRTETRVGIYILSNGEGRARGRDSEATGRSRRKRQIYGYDGRFSIFGKDFLLNYPFS  
 5 TSVKLSTGCTGLVAEKHVLAAHClHDGKTYVKGTQKLRVGF  
 LKPKYKDAGGDNSSSAMPDKMKFQWIRVKR  
 THVPKGWIKGNANDIGMDYD  
 YDLYQQCDAQPGASGSGVYVRMWKR  
 PQQKWERKIIIGIFSGHQWVDMNGSPQDFNVAVRITPLKYAQICYWI  
 KGN  
 YLDCREG

**10 SEQ ID NO:19 Human NK4 nucleic acid sequence**

HUM241786 accession:BC018782 CDS:59..625

GACTGTCTCAGTGGAGCTGGGT  
 CATCTCAGGCCTGGCTCCTGA  
 ACTTTGGCCGCCATGTGCTTCC  
 CGAAGGTCTCTCTGATGACATGA  
 AGAAAGCTGAAGGCCGA  
 ATGCACCAGGCCATAGAAAGATTTAT  
 GATAAAATGCAAATGCAGAATC  
 AGGACGTGGACAGGTGATGTC  
 GAGCCTGGCAGAGCTGGAGGACGA  
 15 CTTCAAAGAGGGCTACCTGGAGACAGTGGCGGCTTATTATGAGGAGCAGCACCCAGAGCTCA  
 TCTC  
 TACTTGAAAAAGAAAGAGATGGATTACGGTGCCGAGGCAACAGATCCCCTG  
 TCCGGATGTTGAGGAT  
 CCCGCAACCGAGGAGCCTGGGAGAGCTTTGT  
 GACAAGGT  
 CATGAGATGGTCCAGGCCATGCTGCA  
 GCGGCTGCAGACCTGGTGGC  
 ACGGGTTCTGGCCTGGTGA  
 AGGAGAAGGTGGTGGCCCTGGTCCATG  
 CAGTGCAGGCCCTCTGGAAACAG  
 TTCCAGGAGCTCTGCTCTGCTCTG  
 20 TTCCAGTCTACGGAGCCCCACGGGGGACAAGGAGGAGCTG  
 ACACCCAGAAGTGCTCTGAACCCCA  
 ATCCTCAAATGAAGATACTGACACCAC  
 CCTTGCCTCCCCGT  
 ACCCGC  
 ACCCACCCTGACCCCTC  
 CCTCAGCTGCTGTG  
 CCCCCGCCCTCTCCCG  
 CACTCAGTCCCCCTGC  
 CTGGCGTTCC  
 CTGCCGCAGC  
 TCTGACCTGGT  
 GCTGTC  
 GCCCTGGCATCTTA  
 AAAACCTG  
 TTACTTCC  
 CCTGGAAAA  
 AAAAAAAA  
 25 AAAAAAAA

25

**SEQ ID NO:20 Human NK4 polypeptide sequence**

protein\_id:gi17511868

MCFPKVLSDDMKKLKARMHQAIERFYDKMQNAESGRGQVMSSL  
 AELEDDFKEGYLETVAAYYEEQHPE  
 LTP  
 LLEKERDGLRCRGNRSPVPDV  
 EDPATEEP  
 GESFC  
 DKVMRWFQAM  
 LQRLQTWWHGVL  
 AWVKE  
 KVVA  
 30 LVHAVQALWKQFQSFCCSL  
 SELFMSSFQSYGAP  
 PRGDKEELTP  
 QKCSE  
 PQSSK

**SEQ ID NO:21 Human Protein C inhibitor nucleic acid sequence**

Nucleotide sequence HUM207659 accession:BC008915 CDS:140..1360

GGCACGAGGCTCAGCTACGGACACATT  
 CAGGTATCCAAGGCAGCAGAGGTGAGTGGGT  
 CCCCCGAG  
 35 CTCTGTGACCTTATGCTCCACACTAA  
 CTCTGGCAGAGCCTCCGTT  
 CCTCATAGAAC  
 AAGAACAGCC  
 ACCATGCAGCT  
 TCCCTCTTGTGC  
 CTGGT  
 GCTCTCAG  
 CCCTCAGGGGG  
 CCTCC  
 ACCGCCA  
 CCACCCCCGG  
 GAGATGA  
 AGAACAGAG  
 AGTC  
 GAGGAC  
 CCT  
 CATGTAG  
 GTGCCACGGT  
 GGCCCC  
 CAGCAGCA

GAAGGGACTTACCTTGACCTACAGGGCCTGGCTCCGCTGCCAGGCCAGAACATCTTCTTC  
 TCCCCTGTGAGCATCTCCATGAGCCTGCCATGCTCTCCCTGGGGCTGGTCCAGCACAAAGATGCA  
 GATCCTGGAGGCCTGGCCTAACCTCCAGAAAAGCTCAGAGAAGGAGCTGCACAGAGGCTTCAGC  
 5 AGCTCCTTCAGGAACCTAACCCAGAGATGGCTCCAGCTGAGCCTCGGAATGCCCTTTCACC  
 GACCTGGTGGTAGACCTGCAGGACACCTCGTAAGTGCATGAAGACGCTGTACCTGGCAGACACTT  
 CCCACCAACTTAAAGGACTCTGCAGGGCCATGAAGCAGATCAATGATTATGTGGCAAAGCAAACGA  
 AGGGCAAGATTGTGGACTTGCTTAAGAACCTCGATAGCAATGCGGTCGTATCATGGTGAATTACATC  
 TTCTTAAAGCTAAGTGGAGACAAGCTAACCAACAAAGGCACCCAAAGAGCAAGACTTACGTGAC  
 CTCGGAGACTGTGGTGCAGGGTACCCATGATGAGCCCGAGGATCAGTATCACTACCTCCTGGACCGGA  
 10 ACCTCTCCTGCAGGGTGGTGGGGTCCCCTACCAAGGCAATGCCACGGCTTGTTCATTCTCCCCAGT  
 GAGGGAAAGATGCAGCAGGTGGAGAATGGACTGAGTGAGAAAACGCTGAGGAAGTGGCTAAGATGTT  
 CAAAAAGAGGCAGCTCGAGCTTACCTCCAAATTCTCATTGAGGGCTCCTATCAGCTGGAGAAAG  
 TCCTCCCCAGTCTGGGATCAGTAACGTCTCACCTCCATGCTATGTCCGGCATCAGCAACCAC  
 15 TCAAATATCCAGGTGTGAGATGGTGCACAAAGCTGTGGTGAGGTGGACGAGTCGGAACCCAGAGC  
 AGCGGCAGCCACGGGACAATCTCACTTCAGGTGGCCCGCTGAACCTCAGAGGCTAGTGTCA  
 ACAGGCCCTTCTGATGTTATTGTGGATAACAAACATCCTCTTGGCAAAGTGAACCGCCCCATGA  
 GGTGGGGCTTCTCTGAAATCTACAGGCTCAGGGTGGAGATGAAGGGGCTATGCTATGGCCCATC  
 TGTATGCTGGTAGCTAGTGAATTACACAGGTTAGTGACTAATGAGGCATTACAAATAATTACTC  
 TATGATGATTGCTTCCACCCACAGCAACTGCAACATACAGGTGCCTGGGGAAATGTGGAGAACATTCA  
 20 ATCTGCCGTCACTATTCAATGAAGATTAGCACTGAGATCCAGAGAGGCTGGATGACTTGCTCAA  
 GTTCACCAGCATGGTAGTGGCAAAGAGAGGTCCAGAGTCCTGGCCCTGATGCCAGCTCAGGCCAC  
 AAAGCTCAGTAGGAGGGATGTCAGTGGATGAGGGCACCAGGAAGCACAGGTCCAAGGCTGGTCCC  
 AACACTTATCAGCAGCAACAATGTCAGTCATTGCTGTGACATTGGGACAACACTTCCCTCTGGACCTCA  
 TGGGGCTACTGTTCACTGCTGTGACCCAGAACACTCCTGAAATCATATAGAAAGG  
 25 GTTTCCCTCTGTATAACAAGGATCAGATTCTGCTGTGACCCAGAACACTCCTGAAATCATATAGAAAGG  
 CTGGGGTGGGCCCTGTCATTGTTGATTCAATACACTCAAGTGCATTGCTGTGACATTGGGACAACACTTCCCTCTGGACCTCA  
 CATCTGGATATCAAGGTGGAAATGGCCATTAAATGATTGATTATATCATTGTGGATATAGTTATA  
 ATCTGATGGCCTGGCTGGAGTGGAGAAGGGAAGCCTTGCAAATAGTAGAGTGTCAAGTTGAG  
 TGCCAATGACTAACTTTGAATTCTATGTTGGCATTAACAATAAGCATTGGAAACAAAAAA  
 30 AAAAAAAA

**SEQ ID NO:22 HumanProtein C inhibitor polypeptide sequence**

protein\_id:gi14286238

MQLFLLLCLVLLSPQGASLHRHHPREMKRVEDLHVGATVAPSSRRDFTFDLYRALASAAPSQNIFFS  
 35 PVSISMSLAMLSLGAGSSTKMQILEGLGLNLQKSSEKELHRGFQQLLQELNQPRDGFQLSLGNALFTD  
 LVVDLQDTFVSAKTLYLADTFPTNFRDSAGAMQKINDYVAKQTKGKIVDLLKNLDSNAVVIMVNYIF  
 FAKWETSFNHKGTQEQDFYVTSETVVRVPMMSREDQYHYLLDRNLSCRVVGVPYQGNATALFILPSE

GKMQQVENGLSEKTLRKWLKMFKKRQLELYLPKFSIEGSYQLEKVLPSLGISNVFTSHADLSGISNHS  
NIQVSEMVHKAVVEVDESGTRAAATGTIFTFRSARLNSQRLVFNRPFLMFIVDNNILFLGVNRP

**SEQ ID NO:23 Mouse Protein C inhibitor nucleic acid sequence**

5 accession:NM\_008785 CDS:125..1342  
 CAGAGTCTCTGTGGTCAGTGACATTGCCGGCTGCAATTCCCTCTGCTCTGCTCTGTCCCTTTGTAA  
 GCCAGGTGGACTATAACCTTGACACTCAGCCACCAACGATAACAGAAGCTGTACCACAATGAGGTTCTC  
 CCCATTCTGTGCCTGGTGCTGTTCATCAGCCATGGGTGGCTTCCCGCCGACACTCCCATTCCAAGAA  
 GAAGAAGGCTAAAGAGTCCTCGGTGGTGCTGGGACCTCCAGTAGCAAAGACTTGTCTTCAGAC  
 10 TCTACAGGGCCTTGGCTTCTGAATCCCCTGGTCAGAATGTCTTCTCCCCCTGAGCGTGTCTATG  
 AGTTGGGCATGCTCTCCCTGGGGCTGGCTTGAAGACGAAGACCCAGATCCTAGATGCCCTAGGCCT  
 CAGCCTCCAGCAAGGCCAAGAAGACAAGCTCCACAAGGGCTTCCAACAGCTGCTACAGAGATTCAAGC  
 AGCCTAGTGAATGGCCTGCAGCTGAGCCTGGCAGTGCCCTTTAAAGACCCAGCAGTACATATCCGG  
 GACGACTTCTGAGTGCATGAAGACACTGTACATGTACAGACACTTCTTACCAACTTGGGAAACCC  
 15 TGAAATTGCCAAGAACAGATCAACAACATATGTAGCCAAGCAGACCAAGGGCAAGATTGTAGACTG  
 TCAAGGACCTTGACAGCACCCATGTCATGATAGTGGAAATTACATCTTCAAAGCCAAGTGGCAG  
 ACGGCCTTCAGTGAGACAAACACCCACAAGATGGATTCCATGTGACCCCCAAAAAGACCATAACGGGT  
 GCCCATGATGAACCGTGAAGATGAGTATTCTACTACCTGGACCAAAACATCTCCTGCACGGTGGTGG  
 GGATCCCTTATCAAGGCAATGCCATTGCTCTATTCTCCCCAGCGAGGGCAAGATGAAGCAGGTG  
 20 GAGGATGGCCTGGACGAGAGAACATTGAGGAACATTGGCTTAAGATGTTACCAAGAGACGCTTAGATCT  
 TTACCTCCCCAAGTTCTCCATTGAGGCTACCTACAAACTGGAAAATGTCTCCCCAAGCTGGCATCC  
 AGGACGTCTCACCAACCCATGCTGACTTGTCTGGCATTACTGACCATAACCAATATCAAGTTGTCTGAG  
 ATGGTGCACAAATCCATGATGGAGGTTGAAGAGTCAGGAACCACAGCAGCTGCCATCACAGGAGCCAT  
 CTTCACATTCAAGATCTGCTCGGCCGAGCTCCCTGAAGATAGAATTACCAAGACCCCTTCTGCTGACCC  
 25 TTATGGAGGATTCACATATACTTTCTGGCAAGGTGACCCGGCCCTGAGGTGGACTCCTCTGAA  
 ACTCATGGCTTACACAGAGGGAGCCAGGCATGTTCTCAGCCCATCCATCTTGGTTGCTAGTGAT  
 TTTCATAGAGTAGGTTGACTAGTAAGGTGTTGAAGTGAATAGATCAATATCCCTATGGCTGCTCCT  
 TGATCACAGTTGCAAGATGTGTTCTGCTCTCATCCTTCATCATTGACACTGACCCCTAAAGAGGTAC  
 TTGATTCTCCAAGGTCTACCATCTTATATGTGACATAGTCGTGCCAAGACTCCAGGCCTCTCCTGC  
 30 TCAGTCTAGCTACAAAGTCCCTAACAGAGACTGGTCCGCATAGATAAAAGGTCAACCAGGAAGCACAGT  
 CATAAACCTGGTCCCCCACCTATCAGCACCAACTCTCAGTTCATCAGGGTTGAGAAAGGACACTGGAT  
 TGAGCTTCTGAAATGCCGCTGGTAGCTCTGTAACATGGGAATGGAATGTCCCTCTGGGCTGTGT  
 TTCTTGTGTGTCAGTGAAGGAACAGAACTAGAAAGCCATTCAACCTGTGAGGATTCTTACACAG  
 GGGTTGGCTCTCTATGTTAGCCAGGGTTGCTGTAAGAACTCAGGTTAGTAGCATACTTGAGTAATCA  
 35 GGGTCAAAGTGTCTCACCAACAAAGGTATCAAATCACAAGCCTGGCTGGCTGAGGTGGTGGTA  
 GAGGTCTGTGGTAAACAGCAGAACTTAGGGTGCAGACTGACTTATTGATTCCGCCTGGCTGGCAAT  
 AAAGGCATTTGCAAAG

**SEQ ID NO:24 Mouse Protein C inhibitor polypeptide sequence**

protein\_id:6679215

MRFFPILCLVLFISHGVASRRSHSKKKAKESSVGAVGPPSSKDFAFRLYRALASESPGQNVFFSPL  
 5 SVSMSLGMILSLGAGLKTKTQILDGLGLSLQQQEDKLHKGFQQLLQRFRQPSDGLQLSLGSALFKDPA  
 VHIRDDFLSAMKTLYMSDTFSTNFGNPEIAKKQINNYVAKQTKGKIVDLIKDLDSHVMIIVNYIFFK  
 AKWQTAFSETNTHKMDFHVTPKKTIRVPMMNREDEYSYLDQNISCTVVGIPYQGNAIALFILPSEGK  
 MKQVEDGLDERTLRNWLKMFTKRRLDLYLPKFSIEATYKLENVLPKLGIQDVFTHADLSGITDHTNI  
 KLSEMVHKSMMEVEESGTTAAAITGAIFTFRSARPSSLKIEFTRPFLLTLMEDSHILFVGKVTRP

10

**SEQ ID NO:25 Rat Protein C nucleic acid sequence**

accession: NM\_022957

CDS:48..1268

TACCCGTCAAGACACTCAGCCACCACAATACAGAAATGGCAGCCACAATGAGGTTCTTCCCCATTCTG  
 TGCCTGGTGCTGTTCTCAGTCACGGGGTGGCTTCCGCCAACGCTCCCATTCTAAGGAGAAGAAGAA  
 15 GTCTAAAGAGTCCTCTGTGGGTGCTGTGGGACTTCCGAAGCAGAGACTTTGCCTTCAGGCTCTACA  
 GGGCCTTGGCTTCTGAAGCCCCTGGTCAGAATGTTTCTTCTCCCCATGAGCGTGTATGAGCTTG  
 GGTATGCTCTCCCTGGGGTCTGGCTTGAAGACTAAGGCACAGATCCTGGAGGGCCTGGCCTCAGCCT  
 CCAGCAAGGCCAAGAGGATATGCTCCACAAGGGCTCCAACAGCTGCTGCAGCAGTTGCCAGCCTA  
 GTGATGCCCTCCAGCTGAGCCTAGGCAGTGCCCTTTACAGACCCAGCAGTACATATCCGGGACCAC  
 20 TTTCTGAGTGCCATGAAGACATTGTACATGTAGACATGTTCTACCAACTTGGGACCCCTGAAAG  
 TGCTAAGAACGAGATCAATGACTATGTAGCCAAGAACGCAAGATTGTAGACTTGATCAAGG  
 ATCTTGACAGCACCCACGTATGGTGGTAAATTATATCTTCTCAAAGCCAAGTGGCAGACGGCC  
 TTCAGTAGCACCAACACCCACAAGATGGATTCCATGTGACTCCAAAGAACGACATTAGGTGCCAAT  
 GATGAACCGCGAAGATATATATTCCCTACATCCTCGACCAAAACATTCCCTGCACGGTGGTGGGATCC  
 25 CTTACCAAGGCAATACATTGCTCTGTTATTCTTCCCAGCGAGGGCAAGATGAAGCGGGTAGAGGAT  
 GCCCTGGATGAGAGAACGTTGAGGAACGGCTTAAGATGTTACAAAGAGACAGTTAGATCTTACCT  
 CCCCAAGTCTCCATTGAGGGTACCTATAAACTGGAAAAAAATCTCCCCAAGCTGGCATCCAGGACA  
 TCTTCACCACCCATGCTGACTTGTCTGGCCTACTGACCACACCAATATTAAGTGTCTGAGATGGTA  
 CACAAATCCATGGTGGAGGTAGACGAGTCAGGAACAAACAGCAGCTGCCACAGGAATTCTCTTCAC  
 30 ACTCAGATCTGCTCGACCGAGCTCTGAAGGTAGAATTCAACCAGACCCCTTCTGGTGGTCATTATGG  
 ATGGTACAAACCTCTATTTCATTGGCAAAGTGATTCAAGCCCTGAGGTGGACTCCTTCTGAAACTCAC  
 AGGCCTATAACAGAACGGAGCCAGGCATATTCTACAGACTATCCACCTCTGGTAGCTAGTGATTTCAT  
 AGAGTAGATTGACTAATAAGGTGTTGTAAGTGAGAAGATCAATATCTGGTGGAGCTCCTTGATCA  
 CAGTTGCAAGATGTGCTCTGCTCTATTGACACTGATGCTAAAGAGGCCACTTGATT  
 35 GCTCCAAGGTCTATCACCATATTGTGACATAGTTATGCCAGACCTCAGACCTCTTGCTCAGTC  
 TAGCGCTACAAAGCTCTGCAAGAGGCTGGTCAGCATAGATAAAGGCCACCAGGAAGCACGGTTATAAA  
 CCTGGTCTCACACCATATTGACACCAGGTTGAGAAAGGACATTGGACTGAGATTCTGAAATGGTCTA

5 ATAGCTATGTAACTCGAAAATGAAAATGTTCCCTCTGCCTGTGTTCTTGAGTATCCAGTCAATGA  
 TCAGAACAGTAGCCCATTCACTGATCTGTGAAATTGTTACACAAGGGTTGGCTACTCTATGCTAGCC  
 ATGGTTGATGTAAGAACTCAGTTAACAGCATACTTGAGCAATCAGGGTAAAAGTGTCCACCATGAA  
 CAAAGGTATCAAATCACAAACCTGGCTGACTGCAGTTGGATGTTAGAGGTCTGTTGAAATAGTAGAA  
 5 CATAGATTGCAGACTGACTTGTGATTCTACTTGGCAATAGCAATAAAGCACTTGCAAAG

**SEQ ID NO:26 Rat Protein C inhibitor polypeptide sequence**

protein\_id: 12621138 CDS:171..1265  
 10 MRFFPILCLVLFFSHGVASRQRSHSKEKKSKESVGAVGTSRSRDFAFRLYRALASEAPGQNVFFSP  
 MSVMSMLGMLSLGSGLKTKAQILEGLGLSLQQGQEDMLHKGFQQLLQQFSQPSDGLQLSLGSALFTDP  
 AVHIRDHFLSAMKTLYMSDMFSTNFGNPESAKQINDYVAKKTNGKIVDLIKLDSTHVMVVNYIFF  
 KAKWQTAFSSTNTHKMDFHVTPKKTIQVPMMNREDIYSYILDQNISETVVGIPYQGNTFALFILPSEG  
 KMKRVEDGLDERTLRNWLKMFTKRQLDLYLPKFSIEGTYKLEKILPKLGIQDIFTTHADLSGLTDHTN  
 IKLSEMVHKSMVEVDESGTAAASTGILFTLRSARPSSLKVEFTRPFLVVIMDGTNLYFIGKVIQP

15

**SEQ ID NO:27 Human MAST205b nucleic acid sequence**

Nucleotide sequence Novel Variant MAST205b CDS:1-5073

ATGTTTCACCCACATCTGCTCCAGCCCTTCCCTCACTAAAGTCCCATTAGTGCTGATTGTGCTTT  
 GGCTACTTCTCCTTGCATTTCTGAACCCACAGGAGCCACAGCAGTCCTGGCACTCCTGTTCCA  
 20 GCCGCCACTGCCGTGGAGTTGTCGGACAAGTAACCGCAAGAGCTTGATTGTGACCTCTAGCACATCA  
 CCTACACTACCACGCCACACTCACCACCTCCATGCCACACAGGTAACAGTCCTTGGACAGCCCCCG  
 GAATTCTCTCAAATGCACCTGCTCATTTCCTTGTCCCTGCCGTAGCCATAGCCACAGAGCTG  
 ACAGGACTGATGGCGCGCTGGCTTGGCCTTTGCCCTCTCAGGATATGGAACTAACACTCCT  
 AGCTCCACTGTCTCATCATGCTCCTCACAGGAAAAGCTGCATCAGTGCTTTCCAGCCTACAGC  
 25 TGATGAGCTGCACTTTGACGAAGCATTCAAGCACAGAGAGCGTACAGATGAGGAAGGACGGCAGT  
 CCCAGCCATGCCCTCGCTCCGGAGCCTCAGTCCGGACGATCCCCAGTATCCTTGACAGTGAA  
 ATAATAATGATGAATCATGTTACAAAGAAAGATTCCAAAGGCCACCGCACAAATGGAAGAGCGACT  
 AGCAGAGTTATTCCTCCAACACTCCAGACAGCGTCTGCCCTGGCAGATGGAGCCCTGAGCTTA  
 TTCATCATCAGGTGATTGAGATGGCCGAGACTGCCTGGATAATCTGGAGTGGCCTCATTACATCA  
 30 CAATACTCTACGAACCTCAAGAGAATTGGAGAAACTTTACAAGATGCTCATGAGCGCTCAGAGAG  
 CTCAGAAGTGGCTTGTGATGCAGCTGGTAAAAAGCTGATGATTATCATTGCCGCCAGCACGTC  
 TCCTGGAATGCCCTGGAGTTGACCTGAAAGAGTTCTACCCGCTACATCGTAGCCAGCTGGCCTCACCCGG  
 AAAGAGGGACAAGGGATTAAATGTGACATTCCCGCTACATCGTAGCCAGCTGGCCTCACCCGG  
 TCCCTAGAAGAAATGGCCAGTTGAGCAGCTGTGACAGTCCTGACACTCCAGAGACAGATGATTCTA  
 35 TTGAGGGCCATGGGCATCTGCCATCTAAAAAGACACCCCTCTGAAGAGGACTTCGAGACCATTAAAG  
 CTCATCAGCAATGGCGCTATGGGCTGTATTCTGGTGGCACAAGTCCACCCGGCAGCGCTTGC  
 CATGAAGAAGATCAACAAGCAGAACCTGATCCTACGGAACAGATCCAGCAGGCCTCGTGGAGCGTG

ACATACTGACTTCGCTGAGAACCCCTTGTGGTCAGCATGTTCTGCTCCTTGATACCAAGCGCAC  
 TTGTGCATGGTATGGAGTACGTTGAAGGGGGAGACTGTGCCACTCTGCTGAAGAATATTGGGCCCT  
 GCCTGTGGACATGGTGCCTATACTTGCAGAAACTGTGCTGGCCCTGGAGTACTTACACAACATG  
 GCATCGTGCACCGTGACCTCAAGCCTGACAACCTCTAATTACATCCATGGGCACATCAAGCTCACG  
 5 GACTTTGGACTGTCCAAAATGGGCCTCATGAGTCTGACAACGAACATTGTATGAGGGTCATATTGAAAA  
 GGATGCCCGGGATTCTGGACAAGCAGGTATGCCGGACCCAGAATACATTGCGCCTGAGGTGATCC  
 TCGGCCAGGGCTATGGGAAGCCAGTGGACTGGTGGCCATGGCATTATCCTGTATGAGTCCTGGT  
 GGCTGCGTCCCTTTGGAGATACTCCGGAGGAGCTTTGGCAGGTGATCAGTGTGAGATTGT  
 GTGGCCTGAGGGTGATGAGGCAGTGCCTTACGACCTCACCTCCAAACTGCTCCACCAGA  
 10 ACCCTCTGGAGAGACTTGGCACAGGCAGTGCCTATGAGGTGAAGCAGCACCCATTCTTACTGGTCTG  
 GACTGGACAGGACTTCTCCGCCAGAAGGCTGAATTATTCTCAGTTGGAGTCAGAGGATGATACTAG  
 CTATTTGACACCCGCTCAGAGCATAACCACATGGACTCGGAGGATGAGGAAGAAGTGTGAGG  
 ATGGCTGCCCTGAGATCCGCCAGTTCTCTCCTGCTCTCCAAGGTTCAACAAGGTGTACAGCAGCATG  
 GAGCGGCTCTCACTGCTCGAGGAGCGCCGGACACCACCCGACCAAGCGCAGCCTGAGTGAGGAGAA  
 15 GGAGGACCATTAGATGGCCTGGCAGGGCTAAAGGCCAGACGGAGCTGGTGATTGGCTCCCTG  
 AGATATTACGGAAGCGGTGTCGGTGTCTGAGTCGTCACAGAGAGTGACTCAAGCCCTCCAATG  
 ACAGTGCAGGCCGCTGCTCAGGCCTCTGGATGCGCTCGTTCCGGAGGCCCTGAGGAGGCCAG  
 CAGCACCCCTCAGGAGGCAACCACAGGAGGGTATATGGTCCTGACACCCCCATCTGGAGAGGGGTAT  
 CTGGCCTGTCACTGAACACTCAGGGAGCAGCCCAAAGCTGGATGAGGAAGCTGTTGGCCGGAGC  
 20 AGTGGTTCCAGTCCAGCTATGGAGACCCGAGGCCGTGGACCTCACAGCTGGTGAGGGAGCCACAGC  
 CAAGGCCATCAGTGAACCTGGCTGTGCGTAGGGCCGCCACGGCTGCTCTGGGACTCAACAGAGA  
 AGCGCACTGCTGCCCTGTCAACAAAGTGTCAAGTCGCCCTCAGCCACAGCCCTCTCACTCCTCATT  
 CCTTCGAACACCACACCTGCTCCCGTTGGCAGCCCCATGTCCCCACATTCTCAGTCGTCACACCC  
 ATCATCCCCGGACTCTCTCCAAGCAGGGACTTCTGCCAGCCCTGGCAGCATGAGGCCCTCCATCA  
 25 TCATCCACCGAGCTGGCAAGAAGTATGGCTCACCCCTGCCGGCATTGCGTCTACATGGGTGACTCC  
 GATGTCTACACCGTGCACCATATGGTGTGGCACGTGGAGGATGGAGGTCCGGCCAGTGAGGCAGGGCT  
 TCGTCAAGGTGACCTCATCACCCATGTCAATGGGAACCTGTGCATGGCTGGTGCACACGGAGGTGG  
 TGGAGCTGATCCTGAAGAGTGGAAACAAGGTGGCATTCAACAACCTCCCTGGAGAACACATCCATT  
 AAAGTGGGCCAGCTCGGAAGGGCAGCTACAAGGCCAAGATGGCCGAAGGAGCAAGAGGAGCCGG  
 30 CAAGGATGGCAAGAAAGCAGAAAAGGAGCTCCCTGTTCCGAAGATCACCAAGCAAGCATCCCTGC  
 TCCACACCAGCCGAGCCTTCTTCCCTTAACCGCTCTTGTCACTAGGGAGAGTGGGCCAGGCTCT  
 CCCACACACGCCAGCCTTCCCCCGATCTCCCACTCAAGGCTACCGGGTGACCCCCGATGCTGT  
 GCATTCAAGTGGAGGGATTCAACAGAGCAGCTCCCCAGCTCCAGCGTGCCTAGTCCCCAGCCG  
 GCTCTGGCACACACGGCCAGCTCCCTCACGGTCTGGCACCCAAGCTCCAACGCCAGTACCGCTCT  
 35 CCACGGCGCAAGTCAGCAGGCAGCATCCACTGTCACCACGGCCACACCCCTCTCCCCACCCCC  
 AACAGCTTCACCTCAGCGGTCCCCATGCCCTGTCTGGCATGTAGCCCAGGCCTTCCCACAAAGC  
 TTCACTTGTCAACCTCCCTGGCAGGCAACTCTCACGGCCAAGAGTGCAGGCCACCCGTTACCA

CTACTCAAGAGGGTGCAGTCGGCTGAGAAACTGGCAGCAGCACTTGCCGCCTCTGAGAAGAAGCTAGC  
 CACTTCTCGAAGCACGCCCTGACCTGCCACTCTGAACTAAAGAAGGAAGTGCCTGCCAGGGAAAG  
 TGAGCCCTCTGGAGGTAGTTGGAGCCAGGAGTGTGCTGTGGCAAGGGGCCCTGCCAGGGAAAGGG  
 GTGCTGCAGCCTGCTCCCTCACGGGCCCTAGGCACCCCTCCGGCAGGACCGAGCCGAACGACGGGAGTC  
 5 GCTGCAGAAGCAAGAACGCCATTCGTGAGGTGGACTCCTCAGAGGACGACACCGAGGAAGGGCCTGAGA  
 ACAGCCAGGGTGCACAGGAGCTGAGCTTGGCACCTCACCCAGAAGTGAGCCAGAGTGTGGCCCTAAA  
 GGAGCAGGAGAGAGTGGGAAGAGGATCCTTCCGTCCAGAGGCCCTAGGAGCCTGGGCCAATGGT  
 CCCAAGCCTATTGACAGGGATCACACTGGGCCTCCCAGAATGGAAAGTCCAGTGGTCCCCACAGGA  
 GGCTCGGGAGCCCACAAGCCATTGAGGAGGCTGCCAGCTCCTCAGCAGGCCAACCTAGGTAG  
 10 TCTGGAGCCACAGACCCATCCCTCCTGAAGGTTGCTGGAAGGCCAGCACCTCCACACCCAGGCAC  
 AACAGCACTTCTCCAGCATTGGACTCACCCACCAGCAGTGTGCTCTCCTCCAGCTCCACCT  
 CTGGGAAGCTGAGCATGTGGCCTGGAAATCCCTTATTGAGGCCAGACAGGGCATCCCAAGCAGA  
 AAGGCAACCAGTGGCAGGTGGCTAGCCAACCTCCAGGATTGGAAAACACAACCTCCAGGCCAGCTAA  
 GAACCTGTCTCCAGGGAGCAGGGGAAGACACAGCCACCTAGTCCCCAGACTGCCCATCCATCTT  
 15 ATGAGGATCCCAGCCAGGGCTGGCTATGGAGTCTGAGTGTGCACAAGCAGTGAAAGAGGATCCAGCC  
 CTGAGCATCACCAAGTGCCTGATGCCTCAGGTGACAGAAGGCAGGACGTTCCATGCCAGGCTGCC  
 CCTCACCCAGAAGTCTGAGCCAGCCTCAGGAGGGCCAAGAACACCAGGGGCATAAAAGCATCGGG  
 ATTTGGCATTGGTCCAGATGAGCTTTAAAGCAAACATAG

20 **SEQ ID NO:28 Human MAST205b novel variant polypeptide sequence**

MFSPTSAPALFLTKVPFSAADCALATSPLAIFLNPRAHSSPGTPCSSRPLPWSCRTRNSRKLIVTSSTS  
 PTLPRPHSPHLHGTGNSPLDPRNFSPNAPAHFSFVPARSHSHRADRTDGRRLWSLASLPSSGYGTNTP  
 SSTVSSSCSSQEKLHQLLFQPTADELHFLTKHFSTESVPDEEGRQSPAMRPRSRSLSPGRSPVSDSE  
 IIMMNHVYKERFPKATAQMEERLAEFISSNTPDSVLPLADGALSFIIHQVIEMARDCLDKSRSGLITS  
 25 QYFYELQENLEKLLQDAHERSESSEVAFVMQLVKKLMIIIARPARLLECLEFDPEEFYHLLEAAEGHA  
 KEGQGIKCDIPRYIVSQLGLTRDPLEEMAQLSSCDSPDTPETDDSIEGHGASLPSKKTPEEDFETIK  
 LIISNGAYGAVFLVRHKSTRQRFAMKKINKQNLLIRNQIQQAFVERDILTFEAENPFVVSMSFCSDTKRH  
 LCMVMEYVEGGDCATLLKNIGALPVDMVRLYFAETVLALEYLHNYGIVHRDLKPDNLLITSMGHIKLT  
 DFGLSKMGLMSLTNLYEGHIEKDAREFLDKQVCGTPEYIAPEVILRQGYGKPVDDWWAMGIILYEFLV  
 30 GCVPFFGDTPEELFGQVISDEIVWPEGDEALPPDAQDLTSKLLHQNPLERLGTGSAYEVKQHPFFTGL  
 DWTGLLRQKAIFIQPLESEDDTSYFDTRSERYHHMDSEDEEEVSEDGCLEIRQFSSCSPRFNKVYSSM  
 ERLSLLEERRTPPTKRSLSSEEKEDHSDGLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPM  
 TVRRRCGGLDAPRFPEGPEEASSTLRRQPQEIGIWVLTTPSGEGVSGPVTEHSGEQRPKLDEEAVGRS  
 SGSSPAMETRGRGTSQLAEGATAKAISDLAVRRARHRLLSGDSTEKRTARPVNKVIKSASATALSLI  
 35 PSEHHTCSPLASPMSPHSQSSNPSSRDSSPSRDFLPALGSMRPPIIIHRAGKKYGFTRAIRVYMGDS  
 DVYTvhHMVWHVEDGGPASEAGLQRQGDLITHVNGEPVHGLVHTEVVELILKSGNKVAISTTPLENTS  
 KVGPARKGSYKAKMARRSKRSRGKDQESRKRSSLFRKITKQASLLHTSRSLSSLNRSLSGGESGPGS

PTHSHSLSPRSPTQGYRVTDAVHSVGGNSSQSSSPSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRS  
PRRKSAGSIPLSPLAHTPSPPPPTASPQRSPSPLSGHVAQAFPTKLHLSPPLGRQLSRPKSAEPPRSP  
LLKRVQSAEKLAALAASEKKLATSRKHSDLPHSELKKELPPREVSPLEVVGARSVLSKGALPGKG  
VLQPAPSRALGTLRQDRAERRESLQKQEAIREVDSSEDDTEEGPENSQGAQELSLAPHPEVSQSVA  
5 GAGESGEEDPFPSRGPRSLGPMVPSLLGITLGPPRMESPSPGPHRRLGSPQAIEEAASSSSAGPNLGQ  
SGATDPIPPEGCWKAQHLHTQALTALSPSTSGLPTSSCSPPSSTSGKLSMWSWKSЛИЕГPDRASPSR  
KATMAGGLANLQDLENTTPAQPKNLSREQGKTQPPSAPRLAHPSYEDPSQGWLWESECAQAVKEDPA  
LSITQVPDASGDRRDVPCRGCPLTQKSEPSLRRGQEПGGHQKHRDLALVPDELLKQT

10 SEQ ID NO:29 Human MAST205 nucleic acid sequence

Nucleotide sequence HUM163512 accession:AB047005 CDS:284..5488

GGCCATGGGCATCTGCCATCTAAAAGACACCCCTCTGAAGAGGGACTTCGAGACCATTAAAGCTCAT  
 CAGCAATGGCGCTATGGGCTGTATTCCTGGTGGCACAAGTCCACCCGGCAGCGCTTGCCATGA  
 AGAAGATCAACAAGCAGAACCTGATCCTACGGAACAGATCCAGCAGGCCTCGTGGAGCGTGACATA  
 CTGACTTCGCTGAGAACCCCTTGTGGTCAGCATGTTCTGCTCCTTGATACCAAGCGCCACTGTG  
 5 CATGGTGTGGAGTACGTTGAAGGGGAGACTGTGCCACTCTGCTGAAGAATATTGGGCCCTGCCTG  
 TGGACATGGTGCCTATACCTTGGCAAACGGCTGGAGTACTTACACAACATGGCATT  
 GTGCACCGTGACCTCAAGCCTGACAACCTCTAATTACATCCATGGGCACATCAAGCTCACGGACTT  
 TGGACTGTCAAAATGGGCCTCATGAGTCTGACAACGAACCTGTATGAGGGTCATATTGAAAAGGATG  
 10 CCCGGGAATTCTGGACAAGCAGGTATGCGGGACCCAGAATACATTGCGCTGAGGTGATCCTGGC  
 CAGGGCTATGGGAAGCCAGTGGACTGGTGGCCATGGCATTATCCTGTATGAGTTCTGGTGGCTG  
 CGTCCCTTTGGAGATACTCCGGAGGAGCTTTGGCAGGTGATCAGTGATGAGATTGTGGC  
 CTGAGGGTGTGAGGCAGTGCCTATGAGGTGAAGCAGCACCCATTCTTACTGGTCTGGACTG  
 15 CTGGAGAGACTTGGCACAGGCAGTGCTGAGGCTGAATTATTCTCAGTTGGAGTCAGAGGATGATACTAGCTATT  
 GACAGGACTTCTCCGCCAGAAGGCTGAATTATTCTCAGTTGGAGTCAGAGGATGATACTAGCTATT  
 TTGACACCCGCTCAGAGCGATACCACACATGGACTCGGAGGATGAGGAAGAAGTGAGTGAGGATGGC  
 20 TGCCTTGAGATCCGCCAGTTCTCTTGCTCTCCAAGGTTCAACAAGGTGTACAGCAGCATGGAGCG  
 GCTCTCACTGCTCGAGGAGCGCCGGACACCACCCCGACCAAGCGCAGCCTGAGTGAGGAGAAGGAGG  
 ACCATTCAAGATGGCTGGCAGGGCTCAAAGGCCAGACCGGAGCTGGTGTAGGGTATTGGCTCCCTGAGATA  
 TTACGGAAGCGGCTGCGGTGCTGAGTCGTCCCACACAGAGAGTGACTCAAGCCCTCCAATGACAGT  
 25 GCGACGCCGCTGCTCAGGCCCTGGATGCGCTCGGTTCCGGAGGGCCCTGAGGAGGCCAGCAGCA  
 CCCTCAGGAGGCAACCACAGGAGGGTATATGGCTCTGACACCCCATCTGGAGAGGGGTATCTGG  
 CCTGTCACTGAACACTCAGGGAGCAGCGCCAAAGCTGGATGAGGAAGCTGGCCAGCAGTGG  
 TTCCAGTCCAGCTATGGAGACCCGAGGCCGTGGACCTCACAGCTGGCTGAGGGAGCCACAGCCAAGG  
 CCATCAGTGACCTGGCTGTGCGTAGGGCCGCCACGGCTGCTCTGGGACTCAACAGAGAAGCGC  
 30 ACTGCTGCCCTGTCAACAAAGTGATCAAGTCCGCTCAGCCACAGCCCTCTCACTCCTATTCTTC  
 GGAACACCACACCTGCTCCCGTTGGCAGCCCATGTCCCCACATTCTCAGTCGTCCAACCCATCAT  
 CCCGGACTCTCTCCAAGCAGGGACTTCTGCCAGCCCTGGCAGCATGAGGCCTCCATCATCATC  
 CACCGAGCTGGCAAGAAGTATGGCTCACCCCTGCCGGCATTGCGTCTACATGGGTGACTCCGATGT  
 CTACACCGTGCACCATATGGTGTGGCACGTGGAGGATGGAGGTCCGGCAGTGAGGCAGGGCTCGTC  
 35 AAGGTGACCTCATACCCATGTCAATGGGAACCTGTGCATGGCTGGTCACACGGAGGTGGTGGAG  
 CTGATCCTGAAGAGTGGAAACAAGGTGGCATTCAACAACCTCCCTGGAGAACACATCCATTAAAGT  
 GGGGCCAGCTCGGAAGGGCAGCTACAAGGCCAAGATGGCCGAAGGAGCAAGAGGAGGCCGGCAAGG  
 ATGGGCAAGAAAGCAGAAAAGGAGCTCCCTGTCAGGCTCACGGCTCTGGAGGATGGCCAGGCTCTCCAC  
 ACCAGCCGAGCCTTCTCCCTTAACCGCTCTGTACAGGGGAGAGTGAGGCTGGCCAGGCTCTCCAC  
 40 ACACAGCCACAGCCTTCCCCCGATCTCCACTCAAGGCTACCGGGTGACCCCGATGCTGTGCATT  
 CAGTGGGAGGGATTCAACAGAGCAGCTCCCCAGCTCCAGCGTGCCTCAGGCTCTCCAGGGCTCTCCAC  
 GGGCACACACGGCCAGCTCCCTCACGGTCTGGCACCCAAGCTCAAACGCCAGTACCGCTCTCCACG

GCGCAAGTCAGCAGGCAGCATCCACTGTCACCACTGGCCACACCCCTCTCCCCACCCCCAACAG  
 CTTCACCTCAGCGGTCCCCATGCCCTGTCTGGCATGTAGCCCAGGCCTTCCCACAAAGCTTCAC  
 TTGTCACCTCCCCTGGGCAGGCAACTCTCACGGCCAAGAGTGCGGAGCCACCCGTTACCAACT  
 CAAGAGGGTGCAGTCGGCTGAGAAACTGGCAGCAGCACTTGCGCCTCTGAGAAGAAGCTAGCCACTT  
 5 CTCGCAAGCACAGCCTTGACCTGCCCACTCTGAACATAAGAAGGAACTGCCGCCAGGGAAGTGAGC  
 CCTCTGGAGGTAGTTGGAGCCAGGAGTGTGCTGTGGCAAGGGGCCCTGCCAGGGAAAGGGGTGCT  
 GCAGCCTGCTCCCTCACGGGCCCTAGGCACCCCTCCGGCAGGACCGAGCCAACGACGGAGTCGCTGC  
 AGAAGCAAGAACGCAATTGAGGTGGACTCCTCAGAGGACGACACCGAGGAAGGGCTGAGAACAGC  
 CAGGGTGCACAGGAGCTGAGCTGGCACCTCACCCAGAAGTGAGCCAGAGTGAGGCCCCCTAAAGGAGC  
 10 AGGAGAGAGTGGGAAGAGGATCCTTCCGTCCAGAGGCCCTAGGAGCCTGGCCAAATGGTCCAA  
 GCCTATTGACAGGGATCACACTGGGCCCTCCAGAATGGAAAGTCCAGTGGTCCCCACAGGAGGCTC  
 GGGAGGCCACAAGCCATTGAGGAGGCTGCCAGCTCCTCAGCAGGCCAACCTAGGTCACTG  
 AGCCACAGACCCCATCCCTCGAAGGTTGCTGGAAGGCCAGCACCTCCACACCCAGGCACAAACAG  
 CACTTCTCCAGCAGTCGGACTCACCCCCACCAGCAGTTGCTCTCCTCCAGCTCACCTCTGG  
 15 AAGCTGAGCATGTGGCCTGAAATCCCTTATTGAGGGCCAGACAGGCATCCCCAAGCAGAAAGGC  
 AACCATGGCAGGTGGCTAGCCAACCTCCAGGATTGGAAACACAACCTCCAGGCCAGCCTAACAG  
 GTCTCCCAGGGAGCAGGGAAAGACACAGCCACCTAGTGCCCCAGACTGGCCATCCATCTTATGAGG  
 ATCCCAGCCAGGGCTGGCTATGGAGTCTGAGTGTGCACAAGCAGTGAAAGAGGATCCAGCCCTGAGC  
 ATCACCCAAGTGCCTGATGCCTCAGGTGACAGAAGGCAGGACGTTCCATGCCAGGCTGCCCTCAC  
 20 CCAGAAGTCTGAGCCAGCCTCAGGAGGGCCAAGAACCAAGGGCCATCAAAGCATCGGGATTG  
 CATTGGTTCCAGATGAGCTTTAAAGCAAACATAGCAGTTGTTGCCATTCTGCACTCAGACCTGT  
 GTAATATATGCTCCTGGAAACCATC

**SEQ ID NO:30 Human MAST205 polypeptide sequence**

25 protein\_id:gi13537204  
 MKRSRCRDRPQPPPPDRREDGVQRAAELSQSLPPRRRAPPRQRLEERTGPAGPEGKEQDVVTGVSP  
 LFRKLSNPDIFSSSTGKVQLQRQLSQDDCKLWRGNLASSLSGKQLLPLSSSVHSSVGQVTWQSSGEASN  
 LVRMRNQSLGQSAPSLTAGLKELSLPRRGFCRTSNRKSILIVTSSTSPTLPRPHSPHLGHTGNSP  
 PRNFSPNAPAHFSFVPARRTDGRRWSLASLPSSGYGTNTPSSTVSSCSSQEKLHQLPFQPTADELF  
 30 LTKHFSTESVPDEEGRQSPAMRPRSRSLSPGRSPVSDSEIIMMNHVYKERFPKATAQMEERLAEFIS  
 SNTPDSVPLADGALSFIIHQVIEMARDCLDKRSRGLITSQYFYELQENLEKLLQDAHERSESEVAF  
 VMQLVKKLMIIIARPALLECLEDPEEFYHLLEAAEGHAKEGQGIKCDIPRYIVSQLGLTRDPLEEM  
 AQLSSCDSPDTPETDDSIEHGASLPSKKTPSEEDFETIKLISNGAYGAVFLVRHKSTRQRFAMKKIN  
 KQNLILRNQIQQAFVERDILTFAENPFVVSFCSDTKRHLCKVMYEYVEGGDCATLLKNIGALPVDMV  
 35 RLYFAETVIALEYLHNYGIVHRDLKPDNLLITSMGHIKLTDGLSKMGLMSLTNTLYEGHIEKDAREF  
 LDKQVCGTPEYIAPEVILRQGYGKPVDDWAMGIILYELVGCVPFFGDTPEELFGQVI  
 SDEIVWPEGDEALPPDAQDLTSKLLHQNP  
 LERLGTGSAYEVKQHPFFTGLDWTGLL  
 RQKAEFI  
 PQLESEDDTSYFDTR

SERYHHMDSEEEVSEDGCLEIRQFSSCSPRFNKVSSMERLSLLEERRTPPTKRSLSSEEKEDHSD  
 GLAGLKGRDRSWIVGSPEILRKRLSVSESSHTESDSSPPMTVRRRCGLLDAPRFPEGPEEASSTLRR  
 QPQEGLWVLTTPSGEGVSGPVTEHSGEQRPKLDEAVGRSSGSSPAMETRGRGTSQLAEGATAKAISD  
 LAVRRARHRLLSGDSTEKRTARPVNKVIKSASATALSLLIPSEHHTCSPLASPMSPHSQSSNPSSRDS  
 5 SPSRDFLPALGSMRPIIIRAGKKYGTIRAIRVYMGSDVYTWHMVHVEDGGPASEAGLRQGDL  
 ITHVNGEPVHGLVHTEVVELILKSGNKVAISTTPLENTSIKVGPKGSYKAKMARRSKRSRGKDQE  
 SRKRSSLFRKITKQASLLHTSRSLSSLNRSLSSESOPGSPTHSHSLSRSPPTQGYRTPDAVHSVGG  
 NSSQSSSPSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRSPRKSAGSIPLSPLAHTPSPPPPTASPO  
 RSPSPLSGHVAQAFPTKLHLSPLGRQLSRPKSAEPPRSPILLKRVQSAEKLAALAASEKKLATSRKH  
 10 SLDLPHSELKKELPPREVSPLEVVGARSVLSGKGALPGKGVLPQAPSRALGTLRQDRAERRESLQKQE  
 AIREVDSSEDDTEEGPENSQGAQELSLAPHPEVQSQSVAPKGAGESGEEDPFPSRGPRSLGPMVPSSLT  
 GITLGPPRMEPSGPHRRLGSPQAIIEAASSSSAGPNLGQSGATDPIPPEGCWKAQHLHTQALTALSP  
 STSGLTPTSSCSPPSSTSGKLSMWSWKSLIEGPDRASPSRKATMAGGLANLQDLETQLQPSLRTCLPG  
 SRGRHSHLVPPDWPIHLMRIPARAGYGSLSVHKQ

15

**SEQ ID NO:31 mouse MAST205 nucleic acid sequence**

accession:NM\_008641

GGACGGCCGAGTTGCCCGAGCCGAGTCAGTCTTGCCCCGCCGGCGAGCGCCGCCGGAGGCAGCTA  
 CTGGAGGAGCGGAGCGGGCTTGGGCACGACAGCAGGGAGCAGGATATGGTTACTGGACTTAGTCC  
 20 TCTGCTCTCAGGAAGCTTAGTAATCCTGACATATTGCACCCACTGGAAAAGTTAAACTCCAGCGAC  
 AACTTAGTCAGGATGACTGTAAGTTACGGAGAGGAAGCCTGGCAAGTTCTCTGTCGGTAAGCAGCTG  
 CTCCCTTGTCAGCAGTGTACACAGCAGTGTGGACAGGTAACCTGGCAGTCTACAGGAGAACATC  
 AACACCTGGTCGAATGAGAAACCAATCCCTGGACAGTCTGCACCTCCCTACGGCTGGCTGAAGG  
 AATTGAGCCTTCCAAGGAGAGGCAGCTTGTGGACAAGTAACCGCAAGAGCTTGATTGTAACCTCC  
 25 AGCACATCACCTACGCTACCAAGGCCACACTCCCCACTCCATGGCACACAGTAACAGCTTGGA  
 CAGCCCCCGGAATTCTCTCAAATGCTCCTGCTCACCTTCCCTGCTGCCGTAGGACTGATG  
 GACGGCGCTGGCTTGGCTCGTGCCTCTCAGGCTATGGAACCAACACTCCTAGCTCTACAGTC  
 TCTTCATCATGTTCCCTACAAGAAAAGCTTCATCAGTTACCTTCCAGCCAACAGCTGATGAACTACA  
 CTTTCTGACGAAGCATTAGCACAGAAAACGTACCAAGATGAGGAGGGACGTCGGTCCCCACGGATGC  
 30 GGGCCCGTTCCCGCAGCCTCAGTCCTGGACGGTCCCAGTTCTTGACAGTGAAATAATAATGATG  
 AATCATGTGTACAAAGAAAGATTCCCCAAGGCCACTGCACAGATGGAAGAGCGACCTAGCCTGACCTT  
 CATTCTCTAACACTCCAGATAGTGTGTTGCCCTGGCAGATGGAGCACTAAGCTTATTCTACATC  
 AGGTGATTGAGATGGCCCGAGACTGCCTGGATAAAATCTGGAGTGGCTCATTACGTACACTATTTC  
 TATGAACCTCAAGAGAATTGGAAAAGCTTCTGCAAGATGCTCACGAACGCTCAGAAAGCTCAGATGT  
 35 AGCCTTGTGATACAGCTGGAAAAAGTTGATGATCATCATTGCTCGCCAGCTCGCCTCTGGAAT  
 GCCTGGAGTTGACCCCTGAAGAATTACCAACCTGTTAGAAGCAGCTGAAGGTCTGCAAAGAGGGAA  
 CATGGAATTAAATGTGACATTCCCCGCTACATCGTTAGCCAGCTGGCCTAACCTGGGATCCCTGGAA

GGAAATGGCCCAGTTGAGCAGCTATGACAGTCCAGATACTCCAGAGACAGATGATTGAGGGTC  
 GTGGGGTATCTCAGCCATCTCAGAAGACCCCTCTGAAGAGGACTTGAAACCATTAAAGCTCATCAGC  
 AATGGCGCCTATGGGCTGTCTTCTGGTGGCACAGTCCACGCCAGCGCTTGCAATGAAGAA  
 GATTAATAAGCAAAACCTAATCCTACGGAACCAGATCCAGCAAGCATTGTGGAACGCGACATACTGA  
 5 CTTTCGCTGAAAACCCTTGTGGTCAGCATGTTCTGCTCCTTGAGACCAAGCGTCACTTATGCATG  
 GTGATGGAATACGTAGAAGGGGAGACTGTGCCACTTGCTCAAGAACATCGGGCCCTACCTGTGGA  
 CATGGTACGCCGTACTTGGGAAACTGTGCTGGCTTGGAATACCTACACAACGACATCGCATCGTC  
 ACCGTGACCTCAAGCCTGACAACCTCTGATTACATCCATGGACACATCAAACACTGACTTTGGA  
 CTTTCTAAAATCGGCCTCATGAGTTGACAACCAACTGTATGAGGGTCATATTGAAAAGGATGCCG  
 10 GGAGTTCCCTAGACAAGCAGGTATGCCGGACCCCGGAATACATAGCACCTGAGGGTATCCTGCGTCAGG  
 GATATGGGAAGCCAGTGGACTGGTGGCCATGGCATAATCCTGTATGAGTTCTGGTGGTGGCTGCGTC  
 CCTTCTTGGAGACACTCCGGAGGAGCTTTGGCAAGTGTGATGAGATCGTGTGGCCCGA  
 AGGTGATGACGCCCTCCCCAGATGCCAAGACCTCACTCCAAACTGCTTCATCAGAACCTAG  
 AAAGACTGGGCACAAGTAGTGCCTATGAAGTGAAGCAGCACCCATTCTCATGGGTGGACTGGACA  
 15 GGACTTCTACGGCAGAAGGCTGAATTATACCTCAGCTGGAGTCAGAGGATGACACCAGCTACTTGA  
 TACCCGTTAGAACGATACCAACCATGTGGACTCTGAGGATGAGGAGGAAGTGAGTGAGGATGGCTGCC  
 TTGAGATTGCCAGTTCTTCTGCTCTCCAGGTTCAAGGTTCAAGGTTACAGTAGCATGGAACGGCTT  
 TCCCTGCTTGAGGAACGCCGACACCACCTCAACCAAGCGCAGCCTCAGTGAGGAAAGGAAGATCA  
 CTCAGACGGCTTGGCAGGACTGAAGGGCGAGATCGCAGCTGGTGATTGGTCCCCGAGATATTAC  
 20 GGAAGCGTTATCTGTGTCCAGTCATCCCACACAGAGAGTGATTGAGTCAGCTCCCAATGACAGTGC  
 CATCGGTGTTCAAGGCCTCCAGATGGCCTCATTGCCCTGAGGAGACTAGTAGCACCCCTCGGAAGCA  
 ACAACAGGAAGGTATATGGGTCTCATCCCCATCTGGAGAGGGTCATCTAGGCTGTTGCCAGCC  
 GACCGTTGGAGAGGCAACTGAAGCTGGATGAGGAGCCTCTGGCCAAAGCAGTCGGTGTGGCCAGCC  
 CTGGAGACACGAGGCCGTGGACCCCTCAGCTAGCTGAGGAAGCTACAGCCAAAGCCATAGCGACCT  
 25 AGCTGTGCGTAGGGCCGTACCGGCTGCTCTGGGACTCTATAGAGAACGCAACACTGCCCTG  
 TCAACAAAGTAATCAAGTCAGCCTCAGCTACGCCCTTCCCTCATTCCAGTCAGAACACCATGCC  
 TGCTCACCATGGCCAGCCCTATGCCCCACATCCCAGTCATCCAATCCATCATCCAGGGACTCTC  
 TCCAAGCAGGGACTTCTGCCAGCCCTGGCAGCTTGAGGAGCCTCCATCATCCACCGAGCTGGCA  
 AGAAGTATGGCTTCACCCCTGGGGCCATTGAGTCTACATGGGTGACACTGATGTCTACACCGTACAC  
 30 CACATGGTGTGGCATGTGGAGGATGGTGGCCAGCCAGTGAAAGCAGGGCTCGTCAGGGTACCTCAT  
 CACCCATGTCAATGGTGAGCCTGTGCATGGCTAGTCCACACAGAACGAGCTGGTGGAGCTGGTCTGAAGA  
 GTGGAAACAAGGTATCAATTCAACAACCTCCCTGGAGAACACAGTCATCAAAGTGGGCCAGCTAGG  
 AAAGGCAGCTATAAGCCAAGATGGCCGAAGGAGCAAACGGAGCAAAGGCAAGGATGGCAAGAAAG  
 CCGAAAAAGAAGCTCCCTATTCCGGAAATCACAAAGCAGGCCCTTGCTCCACACCAGCCGAGCC  
 35 TTTCTCCCTTAACCGCTCCTGTCACTAGGGAGAGTGGTCCAGGCTCTCCACACACAGCCACAGC  
 CTCTCTCCAGATCTCTCCCTCAGGGATACCGGGTAGCCCCAGATGCTGTGCACTCAGTAGGAGGGAA  
 TTCCTCGCAGAGCAGCTCCAGCTCCAGTGCCCCAGTTCTCCTGCTGGCTCTGGACATACACGGC

CCAGCTCTTACGGTCTGGCACCAAGCTCCAACGCCAATACCGCTCACCAACGGCGCAAGTCAGCA  
 GGCAGCATCCACTGTACCAGTTGGCCACACCCCTCCCCACCAGCAACGGCAGCTCACCTCAGCG  
 TTCCCCATCACCTTGTCTGGCATGGTCTCAGTCCTTCTACCAAACCTCACTTGTCTCCTCCGC  
 TAGGTAGGCAGCTCTACGGCCAAGAGTCAGAGCCACCCGCTCTCCCTACTTAAGAGGGTGCAG  
 5 TCTGCTGAGAAGCTGGCGCTGCAGTGGCAGCTGAGAAGAAGTTAGCACCTCCGCAAACATAG  
 TCTTGACCTGCCCCATGGTGAACAAAGAAGGAACGTACACCCAGGAAGCCAGCCCTCTGGAGGTAG  
 TTGGAACCAGAAGTGTGCTATCCGGAAAGGGCACTTCCAGGAAGGGGTACTGCAGCCTGCTCCT  
 TCACGGGCCCTGGGACCCACGGCAGGATCGAGCTGAACGCCGTGAGTCAGTCAAAACAAGAAC  
 AATCCGGGAAGTAGACTCCTCAGAAGATGACACTGATGAGGAGCCTGAGAACAGCCAGGCCACACAGG  
 10 AGCCAAGATTGTCCCCCACCCAGAAGCAAGCCACAATCTACTCCCTAAAGGTCAGGAGAGGGTACA  
 GAAGAGGACACTTCTGCACAGGATCTAAAGAAGCAGGGCCCTGTACTCTCAGGTCTAGTGACAGG  
 GGCCACACTAGGCTCCCCCGAGTAGACGTTCTGGCTCTCCCAAGGAAGGTAGCAGGCCACAAG  
 CCTTGAGGAAGCTACCAACCCCTACAAGTCCCTAGCCTGAGCAGGTCTGGACCCACAAGCCCCACC  
 CCCTCTGAAGGCTGCTGGAAGGCCAGCACCTCACACAGGCACTAACTGCACTTGTCCCAGCTT  
 15 TTCAGAACTTACCCCTACCGGTTGTTCTGCTGCCACCTCCACCTCTGGAAAGCCAGGGACATGGCCT  
 GGAAATTCTTATTGAGGGTCCAGACAGAGCATCCACGAACAAGACATAACAAGGAAGGTGAACCA  
 GCTAACTCCAAGATAACGAATACCACGGTCCAAATCTTCTGAAGAACCTGTCTCTGAGGAGGAGAA  
 GCCACAGCCACCAAGTGTGCTGGCTGACCCATCCGCTTCTTGAGGTCCCCAGCCAGAACTGGCCT  
 GGGAGTCTGAATGTGAACAAATGGAGAAAGAAGAACCATCCCTGAGCATCACCAGTGCCTGATTCC  
 20 TCAGGCGACAGGAGGCAGGACATTCCATGCAGAGCCCACCCCTGAGGCCAGAAACCCGCCAGCCT  
 GCTCTGGAAAAGCCAAGAACTTGGGGCCAGCAAGATCATCAGGACTTAGCAGTACATCAGATGAGC  
 TCTTAAAGCAAACCTAGCACTTGTGCTTCCCTACATTCACCTGTGTAATACACCCCTGGAAAC  
 CA

25 **SEQ ID NO:32 mouse MAST205 polypeptide sequence**

accession:gi6678958

MVTGLSPLLFRKLSNPDI FAPTGVKLQRQLSQDDCKLRRGSLASSLSGKQLLPLSSSVHSSVGQVTW  
 QSTGEASNLVRMRNQSLGQSAPSLTAGLKELSLPRRGSFCRTSNRKSЛИVTSSTSPTLPRPHSPLHGH  
 TGNSPLDSPRNFSNAPAHFSFVPARRTDGRRWSLASLPSSGYGTNTPSSTVSSCSSQEKLHQLPFQ  
 30 PTADELHFLTKHSTENV PDEEGRSPRMRPRSRSLSPGRSPVSDSEIIMMNH VYKERFPKATAQME  
 ERPSLTFIISNTPDSVLPLADGALSFIHHQVIEMARDCLDKSRSGLITSHYFYELQENLEKLLQDAHE  
 RSESSDVAFVIQLVKKLMI IARPARILLEFDPEEFYHLLEAAEGHAKEGHGIKCDIPRYIVSQLG  
 LTRDPLEEMAQLSSYDSPDTPETDDSVEGRGVSQPSQKTPSEEDFETIKLISNGAYGAVFLVRHKSTR  
 QRFAMKKINKQNLILRNQIQQAFVERDILTFAENPFVVS MFCSFETKRHLCMVMEYVEGGDCATLLKN  
 35 I GALPVDMVRLYFAETVLALEYLHN GIVHRDLKPDNLLITSMGHIKLTDGLSKIGLMSLTNLYEG  
 HIEKDAREFLDKQVCGTPEYIAPEVILRQGYGKPVWWAMGIILYEF LVGCVPFFGDTPEELFGQVIS  
 DEIVWPEGDDALPPDAQDLTSKLLHQNPLERLGTSSAYEVKQHPFFMGLDWTGLLRQKAEFI PQESE

DDTSYFDTRSERYHHVDS EDEEEVSEDGCLEIRQFSSCS PRFSKVYSSMERLSLLEERRTPPPTKRS  
 SEEKEDHSDGLLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPMTVRHRCG LPDGPHCPEET  
 SSTPRKQQQEGIWLIPPSGE RSSRPVPERPLERQLKLDEEPPGQSSRCCPALETRGRGTPQLAEEAT  
 AKAISDLAVRRARHRLLSGDSIEKRTTRPVNKVIKSASATLSSLIPSEHHACSPLASPMSPHSQSSN  
 5 PSSRDSSPSRDFLPALGSLRPPIIIHRAGKKYGF TLRAIRVYMGDTDVYT VHHMVWHVEDGGPASEAG  
 LRQGDLITHVNGEPVHGLVHTEVVELVLKSGNKVSISTTPLENTSIKVGPA RKG SYKAKMARRSKRSK  
 GKDGQESRKRSSLFRKITKQASLLHTSRSLSLNRSLSGESGPSPTHSHSLS PRSPPQGYRVAPDA  
 VHSVGGNSSQSSSPSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRSPRRKSAGSIPLSPLAHTPSPPA  
 TAASPQRSPSPLSGHGSQSFPTKLHLS PPLGRQLSRPKSAEPPRSPLLKRVQSAEKLAA ALAAA EKKL  
 10 APSRKHSLDLPHGELKKE LTPREASPLEVVGTRSVLSKGKPLPGKGVLPQAPS RALGTLRQDRAERRE  
 SLQKQEAI REVDSSEDDTDEEPENSQATQEPRLSPHPEASHNLLPKGSGE GT EEDTFLH RD LKKQGPV  
 LSGLVTGATLGS PRVDVPG LSPRKVSRPQAFEEATNPLQVPSL SRSGPTSP TPS EGCWKAQHLHTQAL  
 TALCPSFSELTPTGCSAATSTSGKPGTWSWKFLIEGPDRASTNKTITRKGE PANSQDTNTTVPNLLKN  
 15 LSPEEEKPQPPSVPG LTHPLLEVPSQNWPWESECEQMEKEEP SLSITEVPDSSGDRRQDIPCRAHPLS  
 PETRPSLLWKSQELGGQDHQDLALTSDELLKQT

**SEQ ID NO:33 Human CKLF nucleotide sequence**

HUM224910 accession:AF132818 CDS:537..1910

GCGGCCGCCCTGCGCGCGAAGCTCGTGGCCCGAGAGGGGTGCGGTGGGCCGACGGAGGC GGGCCCT  
 20 GGCTGCCTCTCTCCCTGCTCATAGGCTGGCCGCTCAGGCCCTGGCCGGCTCGGGGCTCGGGATT CGC  
 GCGCGCGCTGCCAATCAGCGATCGGGCCCCGCCCGGGAGTTGGGTGAAATAGAGGCGGGCGTCA  
 AGTGT CAGTAGTCGCGGGCAGGTACGTGCGCTCGGGTCTCTCGCGGAGGTGGCGCCGGAGGACGTTGG  
 CGGGCTCCGGAGAGCCTGAGAGCACGGTGGGGCGGGAGAAAGTGGCCGCCGGAGGACGTTGG  
 CGTTTACGTGTGGAAAGAGCGGAAGAGTTTGCTTTCGTGC CGCCTTCGAAA ACTGCCTGCCGTGT  
 25 CTGAGGAGTCCACCCGAAACCTCCCCTCCCTCCGCCGGCAGCCCCCGCTGAGCTGCCGACCCAAAGCC  
 AGCGTGGCGAGGTGGAAAGTGC GCCC GACCCCGCCTGGAGCTGC GCCCCCGAGTGCCATGGCTAC  
 AAGGGTGTGAGCATGAGCGCCCGCCTGGGACCCGTGCCCGAGCCGCCGGCGCAGGACGAGCCGG  
 TGTTCGCGCAGCTCAAGCCGGTGTGGCGCCGCAATCCGGCCCGACGCCGGCTCTCCCGGC  
 GAGGAGCTGAAGCACCGCACCACCGCCCGCAGGCGCAGCCCGCAGGCCCGCAGGCCGG  
 30 CCAGCCGCCGCCACCGCCGGCTGCCCTCAGAGGACCTGGTCAAACAAGATGTGAAATGGAGA  
 AGTATCTGACACCTCAGCTTCCAGTTCTATAATTCCAGAGCATAAAAGTATAGACGAGACAGT  
 GCCTCAGTCGTAGACCAGTTCTCACTGACACTGAAGGGTTACCTTACAGTATCAACATGAACGTCTT  
 CCTCCCTGACATCACTCACCTGAGAACTGGCCTCTACAAATCCAGAGACCGTGC GTAACACACATCA  
 AGACAGAACCTGTTGCCATTTCAGCCACCAAGAGTGAAACGACTGCCCTCCTCCGGCCCCGACCCAG  
 35 GCCCTCCCTGAGTTCACCAAGTATATT CAGCTCACACCAGACCGCAGCTCCAGAGGTGAACAATATTT  
 CATCAAACAAGAACCTCCCTACACCAGATCTCATCTTCTGTCCCTACCCAGCAGGGCCACCTGTACC  
 AGCTACTGAATA CACCGGATCTAGATATGCCAGTTCTACAAATCAGACAGCAGCAATGGACACTCTT

AATTTCTATGTCAGCTGCCATGGCAGGCCTAACACACACACACCTCTGCTGTTCCGCAGACTGCAGT  
 GAAACAATTCCAGGGCATGCCCTTGACATACACAATGCCAAGTCAGTTCTTCCACAACAGGCCA  
 CTTACTTCCCCGTACCAAGCTCAGAGCCTGGAAGTCCAGATAGACAAGCAGAGATGCTCCAG  
 AATTTAACCCACCTCCATCCTATGCTGCTACAATTGCTTCAAACGGCAATTACAATCAAATTT  
 5 ACCCACCACCTGCCAGTTAACTCACAAAACATCCAACCTGTCAGATACAATAGAAGGAGTAACCCG  
 ATTTGGAGAAACGACGCATCCACTACTGCGATTACCCCTGGTGACAAAAGTTATACCAAGTCTCT  
 CATTAAAAGCTCACCTGAGGACTCACACTGGTGAAAAGCCATACAAGTGTACCTGGGAAGGCTGCGA  
 CTGGAGGTCGCGCGATGGATGAGCTGACCCGCCACTACCGGAAGCACACAGGCGCAAGCCCTCC  
 AGTGCGGGGTGTGCAACCGCAGCTCTCGCGCTTGACCCCTGGCCTGCATATGAAGAGGCACAG  
 10 AACTGAGCACTGCCCGTGTGACCCGTTCCAGGTCCCTGGCTCCCTCAAATGACAGACCTAACTATT  
 CCTGTGAAAAACAACAAAAACAAACAAAAGCAAGAAAACCACAACACTAAAACGAAATGTATATTT  
 TGTATATTGAGAAAACAGGGAAATACATTGTATTAATACCAAAAGTGTGTTGGTATTTAAGAATCTGG  
 AATGCTTGTGTAATGTATATGGCTTACTCAAGCAGATCTCATCTCATGACAGGCAGCCACGTCTCA  
 15 ACATGGGTAAGGGTGGGGTGGAGGGAGTGTGTCAGCGTTTACCTAGGCACCACATTAAATG  
 TGACAGTGTTCAGTAAACAAATCAGTGGCAGGCACCAGAAGAAGAATGGATTGTATGTCAAGATTT  
 ACTTGGCATTGAGTAGTTTTCAATAGTAGGTAATTCTTAGAGATACTGATACCTGGCAATTCA  
 CAAATGCCATTGAACAAATGTGTGGTTTTAAAATTATATACTATATGAGTTGCCTATATTGCT  
 ATTCAAAATTGTAAATATGCAAATCAGCTTATAGGTTATTACAAGTTTTAGGATTCTTGG  
 GGAAGAGTCATAATTCTTGTAAAATAACCATGAATACTACAGTTAGGATTGTGTAAGGTACC  
 20 TCTCAACATTACCAAAATCATTCTTAGAGGGAAAGGAATAATCATTCAAATGAACATTAAAAAGCA  
 AATTTCATGCACTGATTAAAATAGGATTATTTAAATACAAAAGGCATTATATGAATTATAAACTG  
 AAGAGCTTAAAGATAGTTACAAATACAAAAGTCAACCTCTTACAATAAGCTAAACGCAATGTCTTT  
 TAAAAAGAGGACTTAGGGTGTGTTTACATATGACAATGTTGCAATTATGATGCAGTTCAAGTA  
 CCAAAACGTTGAATTGATGATGCACTTACATATCGAGATGTTGCTCGCTCGCAGTACTGTTGGTTA  
 25 AATGACAATTATGTGGATTTGATGTAATACACAGTGAGACACAGTAATTATCTAAATTACAGT  
 GCAGTTAGTTAATCTATTAAACTGACTCAGTGTCTGCCCTTAAATATAAATGATATGTTGAAAAT  
 TAAGGAAGCAAATGCTACATATATGCAATATAAAATAGTAATGTGATGCTGATGCTGTTAACCAAAGG  
 GCAGAATAATAAGCAAAATGCCAAAGGGTCTTAAATTGAAATGAAAATTAAATTGTTTAAAA  
 TATTGTTATCTTATTGAGGGTAATATTGTAAGTTTTAGAAGACAATTTCATAACTG  
 30 TAAATTATAAGTTGTTGTTAGAAAAGTAGCTCTTAAAGATGTAATAGATGACAAACGATGTAAT  
 TAATTGTTGTAAGAGGCTTCAAAATGTTATACGTGGAAACACACCTACATGAAAAGCAGAAATCGGTT  
 GCTGTTGCTCTTCCCTTATTGTTGATTGTGGTCAATTCTATGCAAATAATGGAGCAAA  
 CAGCTGTAGTTGAGAATTGAGAGATGAGATGTTATATTAACGACAATTGTTGG  
 AAAATAAAAGTGCCTAAAGAAAAAA

**SEQ ID NO:34 Human CKLF protein sequence**

protein\_id:gi6580835

MATRVLSMSARLGPVPQPAPQDEPVFAQLKPVVLGAANPARDALFPGEELKHAHHRPQAQPAPAQAP  
QPAQPPATGPRLPPEDLVQTRCEMEKYLTPQLPPVPIIPEHKKYRRDSASVVDQFFTDEGLPYSINM  
NVFLPDITHLRTGLYKSQRPCVTHIKTEPVIAIFSHQSETTAPPAPLTQALPEFTSIFSSHQTAAPEVN  
NIFIKQELPTPDHLHSVPTQQGHLYQLLNTPDLDMPSSTNQTAAMDTLNVMSAAMAGLNHTSAVPQ  
TAVKQFQGMPPCTYTMPSQFLPQQATYFPPSPPSSEPGSPDRQAEMLQNLTPPPSYAATIASKLAIHN  
PNLPTTLPVNSQNIQPVRYNRRSNPDLEKRRIHCDYPGCTKVYTKSSHKAHLRTHTGEKPYKCTWE  
GCDWRFARSDELTRHYRKHTGAKPFQCGVCNRSFSRSDHLALHMKRHQN

10

**SEQ ID NO:35 Mouse CKLF nucleotide sequence**

accession:AF079852 CDS:167..1507

15	CCGAGCCCAGGAGCCCCGATCTCGTCCCCGCTTCGTGAGCGTCTGGCTGCCGGCCAGGGGTCCCC CGCCCGGGCCCCCGCCGAGTCCGCCGTCCCGTGCCAGCCGAGCGAGGTGGATCGCGATCGCTCCG
20	TGTCCCGCTCCCGTAATCCCCAGACCGTCCATGCCACGCCGGTGTGACCATGAGCGCCGCCCTGGG ACCACTGCCAGCCGCCGCCGCAGGCCGAGCCGAGCCGTTCGCCAGCTCAAGCCGGTCTGGCG CTGCGAACCCGGCCCGCAGCGGGCCTTCTCCGGAGACGATCTGAAACACGCGCACCACCCCG CCTGCGCCGCCAGCCGCTGGCCCGACTGCCCTGGAGGAGCTGGTCCAGACAAGATGTGAAAT GGAGAAGTATCTGACCCCTCAGCTCCCTCAGTTCGATAATTTCAGAGCATAAAAGTATAGACGAG
25	ACAGTGCCTCAGTGGTAGACCAAGTTCTCACTGACACTGAAGGCATACCTACAGCATCAACATGAAC GTCTTCCTCCCTGACATCACTCACCTGAGAACTGGCCTCTACAAATCCCAGAGACCATGCGTAACACA GATCAAGACAGAACCTGTTACCATTTCAGCCACCAGAGCGAGTCGACGGCCCTCCTCCCTCCGG CCCCCACCCAGGCTCTCCCGAGTTCACTAGTATCTCAGCTCCACCAGACCACAGCGCCACCACAG GAGGTGAACAATATCTCATCAAACAAGAACTTCTATACCAAGATCTCATCTCTGTCCCTCCCA
30	GCAGGGCCACCTGTACCAGCTGTAAATACACCGGATCTAGACATGCCAGTTCGACAAACCAGACGG CAGTAATGGACACCCCTTAATGTTCTATGGCAGGCCCTAACCCACACCCCTGCTGTTCCACAGACG TCAATGAAACAGTCCAGGGCATGCCCTTGACGTACACCAGCCAAGTCAGTTCTCCACAGCA GGCCACTTATTTCCCCGTCAACCACCAAGCTCAGAGCCTGGAAGTCCGATAGACAAGCTGAGATGC TGCAGAATCTCACCCACCTCCGTCTATGCCCTACAATTGCTTCAAACGGCATTACAACCCA
35	AATTACCTGCCACTCTGCCAGTTAACTCGCAACTCTCCACCTGTCAGATACAACAGAAGGAGTAA CCCGGATCTGGAGAAGCGACGTATCCACTTCTGCGATTATAATGGTTGCACAAAGTTATACAAAGT CGTCTCACTAAAAGCTCACCTGAGGACTCATACGGGCCAGAAGCCCTACAAGTGCACCTGGAGGGC TGCAGTGGAGGTTGCCGGTGGATGAGCTGACCCGCCACTACAGGAAGCACACGGCGCCAAGCC GTTCCAGTGCATGGTGTGCCAACGCCAGCTTCTCCGCTCCGACCACCTCGCGCTGCACATGAAGCGCC AAACCGCAGACCTAACTCATAAAAG

**SEQ ID NO:36 Mouse CKLF protein sequence**

accession:gi4336209

MPTRVLTMSARLGPLPQPPAAQAEPVFAQLKPVLGAANPARDAALFSGDDLKHAHHHPPAPPPAAGPR  
 LPSEELVQTRCEMEKYLTQQLPPVPIISEHKYRRDSASVVDQFFTDEGIPYSINMNVFLPDITHLR  
 5 TGLYKSQRPCVTQIKTEPVTFSHQESTAPPPPAPTQALPEFTSIFSSHQTTAPPQEVDNNIFIKQE  
 LPIPDHLVSPSQQGHLYQLLNTPDLDMPSSNTQAVMDTLNVSMAGLNPHPSAVPQTSMKQFQGMPP  
 CTYTMPSQFLPQQATYFPPSPPSSEPGSPDRQAEMLQNLTPPSYAATIASKLAIHNPNLPATLPVNS  
 PTLPPVRYNRRSNPDLEKRRIHFCDYNGCTKVTKSSHLKAHLRTHTGEKPYKCTWEGCDWRFARSDE  
 LTRHYRKHTGAKPFQCMVCQRFSRSRSDHLALHMKRHQN

10

**SEQ ID NO:37 Rat CKLF nucleotide sequence**

accession:NM\_053394

CGGTATTCAGCTCCACCAGACCACAGGCCAGAGGTGAACAATATCTTCATCAAACAAGAACTTCC  
 TATACCAGATCTTCATCTCTCGTCCCTCCAGCAGGGCACCTGTACAGCTGTTGAATACACCTG  
 15 ATCTAGACATGCCAGTTGACAAACCAAGACAGCAGTCATGGACACCCCTTAATGTCTATGGCTGGC  
 CTTAACTCACACCCCTCTGCTGTGCCACAGACGTCCATGAAACAGTCCAGGGCATGCCTCCTGCAC  
 GTACACCATGCCAGTCAGTTCTTCCACAGCAGGCCACCTACTTCCCCATCACCAACCGAGCTCAG  
 AGCCTGGAAGTCCTGATAGACAAGCTGAGATGCTCCAGAAATCTGACCCACCTCCGTCCTATGCTGCT  
 ACAATTGCTTCGAAACTGGCAATTACAATCCAAATTACCTGCCACTCTGCCAGTTAATTGCCAAA  
 20 TATCCAACCTGTCCGATACAACAGAAGGAGTAACCCGGATCTGGAGAAGCGACGCATCCATTCTGTG  
 ATTATGATGGTTGCACAAAGTTATACAAAGTCGTCTCATTAAAAGCTCACCTGAGGACTCATACG  
 GCGAGAAGCCCTACAAGTGCACCTGGGAGGGCTGCGACTGGAGGTTGCCCCGTCGGACGAGCTGAC  
 CCGCCACTACAGGAAGCACACGGGTGCCAAGCCGTTCCAGTGCCTGGTGTGCAACCGCAGCTCTCCC  
 GCTCCGACCACCTGGCGCTGCACATGAAGCGCCACCAGAACTGAGCACTGCGCACAACCGGCTCGACG  
 25 CCTCGCAGTCCGCTGCCATCCTTAAACCGCAGACCTAACTTCATATAAAAAAAAAAAAAA

**SEQ ID NO:38 Rat CKLF protein sequence**

accession:gi16758122

MPSSTNQAVMDTLNVSMAGLNHSHPAVPQTSMKQFQGMPPCTYTMPSQFLPQQATYFPPSPPSSEPG  
 30 SPDRQAEMLQNLTPPSYAATIASKLAIHNPNLPATLPVNSPNIQPVRYNRRSNPDLEKRRIHFCDYD  
 GCTKVTKSSHLKAHLRTHTGEKPYKCTWEGCDWRFARSDELTRHYRKHTGAKPFQCVVCNRFSRSD  
 HLALHMKRHQN